Recent Work on Intrinsic Value

Edited by

Toni Rønnow-Rasmussen and
Michael J. Zimmerman

Springer

RECENT WORK ON INTRINSIC VALUE

LIBRARY OF ETHICS AND APPLIED PHILOSOPHY

VOLUME 17

Managing Editor:

Govert A. den Hartogh, *University of Amsterdam, The Netherlands*

The titles published in this series are listed at the end of this volume.

RECENT WORK ON INTRINSIC VALUE

Edited by

TONI RØNNOW-RASMUSSEN
Lund University,
Lund, Sweden

and

MICHAEL J. ZIMMERMAN
University of North Carolina at Greensboro,
Greensboro, NC, U.S.A.

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3845-3 (HB)
ISBN-13 978-1-4020-3845-7 (HB)
ISBN-10 1-4020-3846-1 (e-book)
ISBN-13 978-1-4020-3846-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

All Rights Reserved
© 2005 Springer
No part of this work may be reproduced, stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording
or otherwise, without written permission from the Publisher, with the exception
of any material supplied specifically for the purpose of being entered
and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS		vii
SOURCES		ix
INTRODUCTION		xiii
NOTE TO READERS		xxxvii
PART I: IDENTIFYING THE CONCEPT OF INTRINSIC VALUE		
1	Roderick M. Chisholm: *Intrinsic Value*	1
2	Eva Bodanszky and Earl Conee: *Isolating Intrinsic Value*	11
3	Roderick M. Chisholm: *Defining Intrinsic Value*	15
4	Noah M. Lemos: *The Concept of Intrinsic Value*	17
5	Jonathan Dancy: *Should We Pass the Buck?*	33
6	Fred Feldman: *Hyperventilating about Intrinsic Value*	45
PART II: DOUBTS ABOUT THE CONCEPT OF INTRINSIC VALUE		
7	Monroe Beardsley: *Intrinsic Value*	61
8	Christine M. Korsgaard: *Two Distinctions in Goodness*	77
9	Shelly Kagan: *Rethinking Intrinsic Value*	97
10	Wlodek Rabinowicz and Toni Rønnow-Rasmussen: *A Distinction in Value: Intrinsic and For Its Own Sake*	115
11	Judith Jarvis Thomson: *The Right and the Good*	131
12	Michael J. Zimmerman: *Defending the Concept of Intrinsic Value*	153
PART III: IDENTIFYING THE CONCEPT OF INTRINSIC VALUE		
13	Roderick M. Chisholm: *Objectives and Intrinsic Value*	171
14	Noah M. Lemos: *The Bearers of Intrinsic Value*	181
15	Michael J. Zimmerman: *Intrinsic Value and Individual Worth*	191
16	Torbjörn Tännsjö: *A Concrete View of Intrinsic Value*	207
17	Wlodek Rabinowicz and Toni Rønnow-Rasmussen: *Tropic of Value*	213

PART IV: THE LOGIC OF INTRINSIC VALUE

18 Roderick M. Chisholm: *The Intrinsic Value in Disjunctive States of Affairs* 229
19 Philip L. Quinn: *Improved Foundations for a Logic of Intrinsic Value* 241
20 Stuart Rachels: *Counterexamples to the Transitivity of 'Better Than'* 249
21 Ken Binmore and Alex Voorhoeve: *Defending Transitivity against Zeno's Paradox* 265
22 Erik Carlson: *Intransitivity without Zeno's Paradox* 273

PART V: THE COMPUTATION OF INTRINSIC VALUE

23 Neil Feit: *The Structure of Higher Goods* 281
24 Gustaf Arrhenius: *Superiority in Value* 291
25 Roderick M. Chisholm: *Organic Unities* 305
26 Noah M. Lemos: *Chisholm's Definition of Organic Unity* 319
27 Jonathan Dancy: *The Particularist's Progress* 325
28 Gilbert H. Harman: *Toward a Theory of Intrinsic Value* 349
29 Erik Carlson: *The Intrinsic Value of Non-Basic States of Affairs* 361
30 Sven Danielsson: *Harman's Equation and Non-Basic Intrinsic Value* 371
31 Fred Feldman: *Basic Intrinsic Value* 379
32 Michael J. Zimmerman: *Virtual Intrinsic Value and the Principle of Organic Unities* 401

BIBLIOGRAPHY 415

ACKNOWLEDGEMENTS

This publication has been made possible by Stiftelsen Elisabeth Rausings minnesfond, and Erik och Gurli Hultengrens fond. We would like to express our deep gratitude to them for their support. Toni Rønnow-Rasmussen's work on this anthology was supported by a generous grant from The Bank of Sweden Tercentenary Foundation, for which he is grateful. In addition, both editors thank all those who kindly gave their permission for reproduction of copyrighted material. Thanks, too, to two anonymous referees for some very helpful suggestions. Finally, many thanks to Dan Egonsson and Wlodek Rabinowicz for their encouragement and support.

SOURCES

The introduction draws extensively from Michael J. Zimmerman, “Intrinsic vs. Extrinsic Value.” In Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), URL = <http://plato.stanford.edu/archives/fall2004/entries/value-intrinsic-extrinsic/>. © 2004 Michael J. Zimmerman. Material used with the permission of the publisher.

Roderick M. Chisholm, “Intrinsic Value.” In A. I. Goldman and J. Kim, ed., *Values and Morals* (Dordrecht: D. Reidel, 1978): 121-30. © 1978 Kluwer Academic Publishers. Reproduced with the permission of the publisher.

Eva Bodanszky and Earl Conee, “Isolating Intrinsic Value.” *Analysis*, 41 (1981): 51-53. © 1981 Eva Bodanszky and Earl Conee. Reproduced with the permission of the authors.

Roderick M. Chisholm, “Defining Intrinsic Value.” *Analysis*, 41 (1981): 99-100. © 1981 Roderick M. Chisholm. Reproduced with the permission of Eleanor Chisholm.

Noah M. Lemos, “The Concept of Intrinsic Value.” From Noah M. Lemos, *Intrinsic Value* (Cambridge: Cambridge University Press, 1994): 3-19. © 1994 Cambridge University Press. Reproduced with the permission of the author and the publisher.

Jonathan Dancy, “Should We Pass the Buck?” In Anthony O’Hear, ed., *Philosophy, the Good, the True and the Beautiful* (Cambridge: Cambridge University Press, 2000): 159-73. © 2000 Royal Institute of Philosophy and Jonathan Dancy. Reproduced with the permission of the Institute and the author.

Fred Feldman, “Hyperventilating about Intrinsic Value.” *Journal of Ethics*, 2 (1998): 339-54. © 1998 Springer. Reproduced with the permission of the publisher.

Monroe Beardsley, “Intrinsic Value.” *Philosophy and Phenomenological Research*, 26 (1965): 1-17. © 1965 *Philosophy and Phenomenological Research*. Reproduced with the permission of the publisher.

Christine M. Korsgaard, “Two Distinctions in Goodness.” *Philosophical Review*, 92 (1983): 169-95. © 1983 Cornell University. Reproduced with the permission of the author and the publisher.

Shelly Kagan, “Rethinking Intrinsic Value.” *Journal of Ethics*, 2 (1998): 277-97. © 1998 Springer. Reproduced with the permission of the publisher.

Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "A Distinction in Value: Intrinsic and For Its Own Sake." *Proceedings of the Aristotelian Society*, 100 (1999): 33-51. © 1999 the Aristotelian Society. Reproduced by courtesy of the Editor of the Aristotelian Society.

Judith Jarvis Thomson, "The Right and the Good." *Journal of Philosophy*, 94 (1997): 273-98. © 1997 *Journal of Philosophy*. Reproduced with the permission of the author and the publisher.

Michael J. Zimmerman, "Defending the Concept of Intrinsic Value." From Michael J. Zimmerman, *The Nature of Intrinsic Value* (Lanham: Rowman and Littlefield, 2001): 15-32. © 2001 Rowman and Littlefield. Reproduced with the permission of the publisher.

Roderick M. Chisholm, "Objectives and Intrinsic Value." In Rudolf Haller, ed., *Jenseits von Sein und Nichtsein* (Graz: Akademische Druck- und Verlagsanstalt, 1972): 261-69. © 1972 Akademische Druck- und Verlagsanstalt Graz/Austria (ADEVA). Reproduced with the permission of the publisher.

Noah M. Lemos, "The Bearers of Intrinsic Value." From Noah M. Lemos, *Intrinsic Value* (Cambridge: Cambridge University Press, 1994): 20-31. © 1994 Cambridge University Press. Reproduced with the permission of the author and the publisher.

Michael J. Zimmerman, "Intrinsic Value and Individual Worth." In Dan Egonsson, Jonas Josefsson, Björn Petersson, and Toni Rønnow-Rasmussen, ed., *Exploring Practical Philosophy* (Aldershot: Ashgate, 2001): 123-38. © 2001 Ashgate Publishing Ltd. Reproduced with the permission of the publisher.

Torbjörn Tännsjö, "A Concrete View of Intrinsic Value." *Journal of Value Inquiry*, 33 (1999): 531-36. © 1999 Springer. Reproduced with the permission of the publisher.

Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "Tropic of Value." *Philosophy and Phenomenological Research*, 66 (2003): 389-403. © 2003 *Philosophy and Phenomenological Research*. Reproduced with the permission of the publisher.

Roderick M. Chisholm, "The Intrinsic Value in Disjunctive States of Affairs." *Noûs*, 9 (1975): 295-308. © 1975 Blackwell. Reproduced with the permission of the publisher.

Philip L. Quinn, "Improved Foundations for a Logic of Intrinsic Value." *Philosophical Studies*, 32 (1977): 73-81. © 1977 Springer. Reproduced with the permission of the publisher.

Stuart Rachels, "Counterexamples to the Transitivity of 'Better Than'." *Australasian Journal of Philosophy*, 76 (1998): 71-83. © 1998 Oxford University Press. Reproduced with the permission of the publisher.

Ken Binmore and Alex Voorhoeve, "Defending Transitivity against Zeno's Paradox." *Philosophy and Public Affairs*, 31 (2003): 272-79. © 2003 Blackwell. Reproduced with the permission of the publisher.

Erik Carlson, "Intransitivity without Zeno's Paradox." © 2004 Erik Carlson. First publication. Printed with the permission of the author.

Neil Feit, “The Structure of Higher Goods.” *Southern Journal of Philosophy*, 39 (2001): 47-57. © 2001 *Southern Journal of Philosophy*. Reproduced with the permission of the publisher.

Gustaf Arrhenius, “Superiority in Value.” *Philosophical Studies*, forthcoming. © 2005 Springer. Reproduced with the permission of the publisher.

Roderick M. Chisholm, “Organic Unities.” From Roderick M. Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986): 69-89. © 1986 Cambridge University Press. Reproduced with the permission of the publisher and Eleanor Chisholm.

Noah M. Lemos, “Chisholm’s Definition of Organic Unity.” From Noah M. Lemos, *Intrinsic Value* (Cambridge: Cambridge University Press, 1994): 196-200. © 1994 Cambridge University Press. Reproduced with the permission of the author and the publisher.

Jonathan Dancy, “The Particularist’s Progress.” In Brad Hooker and Margaret Little, ed., *Moral Particularism* (Oxford: Clarendon Press, 2000): 130-56. © 2000 Clarendon Press. Reproduced with the permission of the publisher.

Gilbert H. Harman, “Toward a Theory of Intrinsic Value.” *Journal of Philosophy*, 64 (1967): 792-804. © 1967 *Journal of Philosophy*. Reproduced with the permission of the author and the publisher.

Erik Carlson, “The Intrinsic Value of Non-Basic States of Affairs.” *Philosophical Studies*, 85 (1997): 95-107. © 1997 Springer. Reproduced with the permission of the publisher.

Sven Danielsson, “Harman’s Equation and Non-Basic Intrinsic Value.” In Lars Lindahl, Paul Needham, and Rysiek Sliwinski, ed., *Not Without Cause* (Uppsala: Uppsala Philosophical Studies 47, 1998): 11-19. © 1998 Sven Danielsson. Reproduced with the permission of the author.

Fred Feldman, “Basic Intrinsic Value.” *Philosophical Studies*, 99 (2000): 319-46. © 2000 Springer. Reproduced with the permission of the publisher.

Michael J. Zimmerman, “Virtual Intrinsic Value and the Principle of Organic Unities.” *Philosophy and Phenomenological Research*, 59 (1999): 653-66. © 1999 *Philosophy and Phenomenological Research*. Reproduced with the permission of the publisher.

INTRODUCTION

Intrinsic value has traditionally been thought to lie at the heart of ethics. Philosophers use a number of terms to refer to such value. The intrinsic value of something is said to be the value that that thing has "in itself," or "for its own sake," or "as such," or "in its own right." Extrinsic value is value that is not intrinsic.

Many philosophers take intrinsic value to be crucial to a variety of moral judgments. For example, according to a fundamental form of so-called consequentialism, whether an action is *morally right or wrong* has exclusively to do with whether its consequences are intrinsically better than those of any other action one can perform under the circumstances.[1] Many other theories also hold that what it is right or wrong to do has at least in part to do with the intrinsic value of the consequences of the actions one can perform. Moreover, if, as is commonly believed, what one is *morally responsible* (praiseworthy or blameworthy) for doing is some function of the rightness or wrongness of what one does, then intrinsic value would seem relevant to judgments about responsibility, too. Intrinsic value is also often taken to be pertinent to judgments about *moral justice* (whether having to do with moral rights or moral desert), insofar as it is good that justice is done and bad that justice is denied, in ways that appear intimately tied to intrinsic value. Finally, it is typically thought that judgments about *moral virtue and vice* also turn on questions of intrinsic value, inasmuch as virtues are good, and vices bad, in ways that appear closely connected to such value.

All four types of moral judgments have been the subject of discussion since the dawn of western philosophy in ancient Greece. The Greeks themselves were especially concerned with questions about virtue and vice, and the concept of intrinsic value may be found at work in their writings and in the writings of moral philosophers ever since. Despite this fact, and rather surprisingly, it is only within the last one hundred years or so that this concept has itself been the subject of sustained scrutiny, and even within this relatively brief period the scrutiny has waxed and waned. At the moment it is intensifying considerably. With this anthology we hope to provide a comprehensive and balanced picture of current

---

[1] See, e.g., G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903 [revised edition 1993]), and *Ethics* (Oxford: Oxford University Press, 1912). Moore's view is discussed by Judith Thomson in her contribution to Part II of this volume.

thinking about intrinsic value, one that will provide a backdrop against which future contributions to the subject may be assessed.

We have divided the anthology into five parts that together cover all the traditional areas of inquiry concerning the concept of intrinsic value. The first part deals with the most fundamental question of all – how exactly the concept is to be understood – while the second discusses a number of doubts that have been raised about the concept. The third part is concerned with what sort or sorts of things could sensibly be said to have intrinsic value, the fourth with questions about the logic of intrinsic value, and the fifth with problems in the computation of intrinsic value.

In this introduction we will set the stage for what follows by providing a background for and brief summaries of the contributions to this volume.

1. IDENTIFYING THE CONCEPT OF INTRINSIC VALUE

As we have said, the concept of intrinsic value features prominently in some of the earliest writings on ethics. For example, in his dialogue *Protagoras*, Plato [430-347 B.C.E.] maintains (through the character of Socrates, modeled after the real Socrates [470-399 B.C.E.], who was Plato's teacher) that, when people condemn pleasure, they do so, not because they take pleasure to be bad as such, but because of the bad consequences they find pleasure often to have. As Socrates says to Protagoras:

> [T]he only reason why these pleasures [of food and drink and sex] seem to you to be evil is...that they result in pain and deprive us of future pleasures.[2]

Socrates concludes that pleasure is in fact good as such and pain bad, regardless of what their consequences may on occasion be.[3] In the *Timaeus*, Plato seems quite pessimistic about these consequences, for he has Timaeus declare pleasure to be "the greatest incitement to evil" and pain to be something that "deters from good."[4] Despite this, he does not appear to relinquish the view that pleasure is good as such and pain bad. (Plato does not think of pleasure as the "highest" good, however. In the *Republic*, Socrates states that there can be no "communion" between "extravagant" pleasure and virtue;[5] and in the *Philebus*, where Philebus argues that pleasure is the highest good, Socrates argues against this, claiming that pleasure is better when accompanied by intelligence.[6])

Many philosophers have followed Plato's lead in declaring pleasure intrinsically good and pain intrinsically bad. Aristotle [384-322 B.C.E.], for example, himself a student of Plato's, says this:

> [I]t is agreed that pain is bad and to be avoided; for some pain is without qualification bad, and other pain is bad because it is in some respect an impediment to us. Now the contrary of that which is to be avoided, *qua*

[2] Plato, *Protagoras*, 353e.
[3] *Ibid.*, 358a.
[4] Plato, *Timaeus*, 69d.
[5] Plato, *Republic*, 402e.
[6] Plato, *Philebus*, 60e.

> something to be avoided and bad, is good. Pleasure, then, is necessarily a good.[7]

Over the course of the more than two thousand years since this was written, this view has been frequently endorsed. Like Plato, Aristotle did not take pleasure and pain to be the only things that are intrinsically good and bad, although some have maintained that this is indeed the case. Jeremy Bentham [1748-1832], for example, makes a very clear statement to this effect:

> Strictly speaking, nothing can be said to be good or bad, but either in itself; which is the case only with pain or pleasure: or on account of its effects; which is the case only with things that are the causes or preventives of pain and pleasure.[8]

This view, often called hedonism, of which Bentham was by no means the first proponent (Epicurus [341-271 B.C.E.] had endorsed a version of the doctrine long before), is still the subject of intense debate today.[9]

It is evident, then, that for a great many years philosophers have unhesitatingly and routinely put the concept of intrinsic value to use. Let us now say something more about this concept in order to give you a better idea of just what it involves. In the opening paragraph, we used several terms to characterize the concept; we said that the intrinsic value of something is the value that it has "in itself," or "for its own sake," or "as such," or "in its own right." The custom has been not to distinguish between the meanings of these terms, but we will see that there is reason to think that there may in fact be more than one concept at issue here. For the moment, though, let us ignore this complication and focus on what it means to say that something is valuable *for its own sake* as opposed to being valuable *for the sake of something else* to which it is related in some way. Perhaps it is easiest to grasp this distinction by way of illustration.

Suppose that someone were to ask you whether it is good to help others in time of need. Unless you suspected some sort of trick, you would answer, "Yes, of course." If this person were to go on to ask you why acting in this way is good, you might say that it is good to help others in time of need simply because it is good that their needs be satisfied. If you were then asked why it is good that people's needs be satisfied, you might be puzzled. You might be inclined to say, "It just is." Or you might accept the legitimacy of the question and say that it is good that people's needs be satisfied because this brings them pleasure. But then, of course, your interlocutor could ask once again, "What's good about that?" Perhaps at this point you would answer, "It just is good that people be pleased," and thus put an end to this line of questioning. Or perhaps you would again seek to explain the fact that it is good that people be pleased in terms of something else that you take to be good. At

---

[7] Aristotle, *Nicomachean Ethics*, 1153b.

[8] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (London: Athlone Press, 1970), p. 11.

[9] This account of hedonism is rough, since hedonists often allow for the intrinsic goodness and badness of things *other* than, but containing, pleasure and pain (things such as lives and worlds, for example).

some point, though, you would have to put an end to the questions, not because you would have grown tired of them (though that is a distinct possibility), but because you would be forced to recognize that, if one thing derives its goodness from some other thing, which derives its goodness from yet a third thing, and so on, there must come a point at which you reach something whose goodness is not derivative in this way, something that "just is" good in its own right, something whose goodness is the source of, and thus explains, the goodness to be found in all the other things that precede it on the list. It is at this point that you will have arrived at intrinsic goodness.[10] That which is intrinsically good is nonderivatively good; it is good for its *own* sake. That which is extrinsically good is derivatively good; it is good, not (insofar as its extrinsic value is concerned) for its own sake, but for the sake of something else that is good and to which it is related in some way. Intrinsic value thus has a certain priority over extrinsic value. The latter is derivative from or reflective of the former and is to be explained in terms of the former. It is for this reason that philosophers have tended to focus on intrinsic value in particular.

The account just given of the distinction between intrinsic and extrinsic value is rough, but it should do as a start. Certain complications must be immediately acknowledged, though. First, there is the possibility, mentioned above, that the terms traditionally used to refer to intrinsic value in fact refer to more than one concept; again, this will be addressed later (in this section and the next). In fact, there are other complications, too. (Sorry!) First, it may not in fact be accurate to say that whatever is intrinsically good is nonderivatively good; some intrinsic value may be derivative. This is a matter we'll discuss in Section 5; again, though, you may safely ignore it for now. Another complication is this. It is almost universally acknowledged among philosophers that all value is "supervenient" on certain nonevaluative features of the thing that has value. Roughly, what this means is that, if something has value, it will have this value in virtue of certain nonevaluative features that it has; its value can be attributed to these features. For example, the value of helping others in time of need might be attributed to the fact that such behavior has the feature of being causally related to certain pleasant experiences induced in those who receive the help. Suppose we accept this and accept also that the experiences in question are intrinsically good. In saying this, we are (barring the complication to be discussed in Section 5) taking the value of the experiences to be nonderivative. Nonetheless, we may well take this value, like all value, to be supervenient on something. In this case, we would probably simply attribute the value of the experiences to their having the feature of being pleasant. This brings out the subtle but important point that the question whether some value is derivative is distinct from the question whether it is supervenient. Even nonderivative value (value that something has in its own right; value that is, in some way, not attributable *to the value* of anything else) is usually understood to be supervenient on certain nonevaluative features of the thing that has value (and thus to be attributable, in a different way, *to these features*).

---

[10] Cf. Aristotle, *op. cit.*, 1094a.

To repeat: whatever is intrinsically good is (barring the complication to be discussed in Section 5) nonderivatively good. It would be a mistake, however, to affirm the converse of this and say that whatever is nonderivatively good is intrinsically good. As "intrinsic value" is traditionally understood, it refers to a *particular way* of being nonderivatively good; there are other ways in which something might be nonderivatively good. For example, suppose that your interlocutor were to ask you whether it is good to eat and drink in moderation and to exercise regularly. Again, you would say, "Yes, of course." If asked why, you would say that this is because such behavior promotes health. If asked what is good about being healthy, you might cite something else whose goodness would explain the value of health, or you might simply say, "Being healthy just is a good way to be." If the latter were your response, you would be indicating that you took health to be nonderivatively good in some way. In what way, though? Well, perhaps you would be thinking of health as intrinsically good. But perhaps not. Suppose that what you meant was that being healthy just is "good for" the person who is healthy (in the sense that it is in each person's interest to be healthy), so that John's being healthy is good for John, Jane's being healthy is good for Jane, and so on. You would thereby be attributing a type of nonderivative interest-value to John's being healthy, and yet it would be perfectly consistent for you to deny that John's being healthy is *intrinsically* good. If John were a villain, you might well deny this. Indeed, you might want to insist that, in light of his villainy, his being healthy is intrinsically *bad*, even though you recognize that his being healthy is good *for him*. If you did say this, you would be indicating that you subscribe to the common view that intrinsic value is nonderivative value of some peculiarly *moral* sort.[11]

Let us now see whether this still rough account of intrinsic value can be made more precise. One of the first writers to concern himself with the question of what exactly is at issue when we ascribe intrinsic value to something was G. E. Moore [1873-1958]. In his book *Principia Ethica*, Moore asks whether the concept of intrinsic value (or, more particularly, the concept of intrinsic goodness, upon which he tended to focus) is analyzable. In raising this question, he has a particular type of analysis in mind, one which consists in "breaking down" a concept into simpler component concepts. (One example of an analysis of this sort is the analysis of the concept of being a vixen in terms of the concepts of being a fox and being female.) His own answer to the question is that the concept of intrinsic goodness is *not* amenable to such analysis.[12] In place of analysis, Moore proposes a certain kind of thought-experiment in order both to come to understand the concept better and to reach a decision about what is intrinsically good. He advises us to consider what things are such that, if they existed by themselves "in absolute isolation," we would judge their existence to be good; in this way, we will be better able to see what

---

[11] Although this is how the term "intrinsic value" is often understood, it has been understood in other ways, too. Sometimes it appears to be used simply as a synonym of "nonderivative value" after all. Also, at one point Moore uses it to refer to any kind of value that supervenes solely on the intrinsic nature of the value-bearer. See G. E. Moore, *Philosophical Studies* (New York: Harcourt, Brace, 1922), p. 260.

[12] Moore, *Principia Ethica*, ch. 1.

really accounts for the value that there is in our world. For example, if such a thought-experiment led you to conclude that all and only pleasure would be good in isolation, and all and only pain bad, you would be a hedonist.[13] Moore himself deems it incredible that anyone, thinking clearly, would reach this conclusion. He says that it involves our saying that a world in which only pleasure existed – a world without any knowledge, love, enjoyment of beauty, or moral qualities – is better than a world that contained all these things but in which there existed slightly less pleasure.[14] Such a view he finds absurd.

Regardless of the merits of this isolation test, it remains unclear exactly why Moore finds the concept of intrinsic goodness to be unanalyzable. At one point he attacks the view that it can be analyzed wholly in terms of "natural" concepts – the view, that is, that we can break down the concept of being intrinsically good into the simpler concepts of being *A*, being *B*, being *C*..., where these component concepts are all purely descriptive rather than evaluative. (One candidate that Moore discusses is this: for something to be intrinsically good is for it to be something that we desire to desire.) He argues that any such analysis is to be rejected, since it will always be intelligible to ask whether (and, presumably, to deny that) it is good that something be *A*, *B*, *C*..., which would not be the case if the analysis were accurate.[15] Even if this argument is successful (a complicated matter about which there is considerable disagreement), it of course does not establish the more general claim that the concept of intrinsic goodness is not analyzable at all, since it leaves open the possibility that this concept is analyzable in terms of other concepts, some or all of which are not "natural" but evaluative. Moore apparently thinks that his objection works just as well where one or more of the component concepts *A*, *B*, *C*... is evaluative; but, again, many dispute the cogency of his argument. Indeed, several philosophers have proposed analyses of just this sort. For example, Roderick Chisholm [1916-1999] argues in Chapter 1 of this volume that Moore's own isolation test in fact provides the basis for an analysis of the concept of intrinsic value. He formulates a view according to which (to put matters roughly) a state of affairs has intrinsic value just in case it is possible that it contain all the value that there is in the world.

Despite the care that Chisholm takes in developing his view, Eva Bodanszky and Earl Conee argue in Chapter 2 that the view is defective; they go on to venture the opinion that the project of analyzing the concept of intrinsic value in terms of the isolation approach is doomed to fail. In Chapter 3 Chisholm responds by accepting that Bodanszky and Conee have shown his particular proposal to be unacceptable, but he denies that the isolation approach is to be abandoned. On the contrary, he proffers a new analysis, one that constitutes a shift from the "ontological isolationism" of his first proposal to a form of "intentional isolationism."[16] Instead

---

[13] Again, this is to put matters only roughly. See n. 9 above.
[14] Moore, *Ethics*, p. 102.
[15] Moore, *Principia Ethica*, pp. 15-16.
[16] Chisholm does not himself use these terms. Noah M. Lemos introduces them in *Intrinsic Value* (Cambridge: Cambridge University Press, 1994), pp. 10-11.

of thinking of intrinsic value in terms of the value that something would have if it *existed* in isolation (an "ontological" matter), Chisholm now proposes to think of it in terms of the kind of attitude it would be appropriate to have if one were to *contemplate* the valuable thing as such, without reference to circumstances or consequences (an "intentional" matter). In Chapter 4 Noah Lemos pursues this proposal, although he doubts whether it is acceptable.

Since both Chisholm and Lemos present their proposals in a clear and straightforward way, there is no need here to go into the details of what they have to say. However, it should be noted that the general idea underlying their proposals is one that has a rich history. Franz Brentano [1838-1917], C. D. Broad [1887-1971], W. D. Ross [1877-1971], and A. C. Ewing [1899-1973], among others, have claimed, in a more or less qualified way, that the concept of intrinsic goodness is analyzable in terms of the worthiness of some attitude.[17] Such an analysis is supported by the mundane observation that, instead of "good," we often use the term "valuable," which itself just means: worthy of being valued. It would thus seem very natural to suppose that for something to be intrinsically good is simply for it to be worthy of being valued for its own sake. ("Worthy" here is usually understood to signify a particular kind of moral worthiness, in keeping with the idea that intrinsic value is a particular kind of moral value. The underlying point is that those who value for its own sake that which is intrinsically good thereby evince a kind of *moral* sensitivity.)

Though undoubtedly attractive, this analysis can be and has been challenged. Brand Blanshard [1892-1987], for example, has claimed that, even if it is necessarily true that whatever is intrinsically good is worthy of being valued for its own sake, and *vice versa*, the proposed analysis of the concept of intrinsic goodness in these terms must be rejected because, if we ask *why* something is worthy of being valued for its own sake, the answer is that this is the case precisely *because* the thing in question is intrinsically good; this answer indicates that the concept of intrinsic goodness is more fundamental than that of the worthiness of being valued, which is inconsistent with analyzing the former in terms of the latter.[18] Ewing and others have resisted this type of argument.[19] In Chapter 4 Lemos entertains both this argument and others for rejecting any attempt to analyze the concept of intrinsic value in the spirit of intentional isolationism. In Chapter 5 Jonathan Dancy gives still further reasons to be skeptical of any such attempt. We will not try to settle the issue here. Whatever the final verdict about such an attempt should be, it is worth noting the following point (one that is stressed by both Lemos and Dancy). Even if all attempts at such an analysis do and must fail, it may nonetheless be necessarily true that whatever is intrinsically good is worthy of being valued for its own sake, and

[17] Franz Brentano, *The Origin of Our Knowledge of Right and Wrong* (London: Routledge and Kegan Paul, 1969), p. 18; C. D. Broad, *Five Types of Ethical Theory* (London: Kegan Paul, Trench, Trubner, 1930), p. 283; W. D. Ross, *Foundations of Ethics* (Oxford: Oxford University Press, 1939), pp. 275-76; A. C. Ewing, *The Definition of Good* (New York: Macmillan, 1948), p. 152.

[18] Brand Blanshard, *Reason and Goodness* (London: Allen and Unwin, 1961), pp. 284-86.

[19] Ewing, *op. cit.*, pp. 157 and 172; cf. Lemos, *op. cit.*, p. 19.

vice versa. If such were the case, it would reveal an important feature of intrinsic value, although there would admittedly remain pressing questions as to just how the concept of intrinsic value is to be understood. In Chapter 6 Fred Feldman pursues some of these questions. He focuses in particular on four "guiding intuitions" about intrinsic value to which many philosophers have subscribed. He finds that two of them are mistaken and that the other two fail to apply uniquely to intrinsic value. The upshot, according to Feldman, is that the concept of intrinsic value unfortunately remains somewhat obscure, so that those (such as himself) who believe that this concept lies at the heart of ethics must continue to search for ways to identify the concept in order to win over those (such as some of the philosophers mentioned in the next section) who have doubts about its usefulness or coherence.

One final cautionary note: it is apparent that some philosophers use the term "intrinsic value" and similar terms to express a concept that is in some ways similar to, but in other ways quite different from, the one that we have just been discussing. Immanuel Kant [1724-1804], for example, is famous for saying:

> It is impossible to conceive anything at all in the world, or even out of it, which can be taken as good without qualification, except a *good will*.[20]

He adds:

> A good will is not good because of what it effects or accomplishes – because of its fitness for attaining some proposed end; it is good through its willing alone – that is, good in itself.[21]

These remarks may seem to suggest that Kant ascribes (positive) intrinsic value only to a good will, declaring the value that anything else may possess merely extrinsic, in the senses of "intrinsic value" and "extrinsic value" discussed above. This suggestion is, if anything, reinforced by what Kant immediately goes on to say:

> Considered in itself [a good will] is to be esteemed beyond comparison as far higher than anything it could ever bring about… Even if…this will [were] entirely lacking in power to carry out its intentions…it would still shine like a jewel for its own sake as something which has full value in itself. Its usefulness…can neither add to, nor subtract from, this value.[22]

Here Kant may seem not only to be invoking the distinction between intrinsic and extrinsic value but also to be in agreement with Brentano *et al.* regarding the characterization of the former in terms of the worthiness of some attitude, namely, esteem. (The term "respect" is often used in place of "esteem" in such contexts.) Nonetheless, it becomes clear on further inspection that in these passages Kant is in fact discussing a concept distinct from that with which this anthology is concerned. A little later on he says that all rational beings, even those that lack a good will, have

---

[20] Immanuel Kant, *Groundwork of the Metaphysic of Morals* (New York: Harper and Row, 1964), p. 61 (Ak. 1).
[21] *Ibid*., p. 62 (Ak. 3).
[22] *Idem*.

"absolute value"; such beings are "ends in themselves" that have a "dignity" or "intrinsic value" that is "above all price."[23] Such talk indicates that Kant believes that the sort of value that good wills in particular, or rational beings in general, possess is infinitely great. But then, if this were understood as a thesis about intrinsic value as we have been understanding this concept, the implication would seem to be that this world is as good as it could be.[24] Yet this is something that Kant explicitly rejects elsewhere. He says:

> What does the highest good consist in? The most perfect world is the highest created good. But the most perfect world involves the happiness of rational creatures and the worthiness of those creatures for such happiness... If the world were full of...rational creatures, who were all well-behaved, and thus worthy of happiness, and they were in the neediest of circumstances, surrounded with sorrow and trouble, they would then have no happiness, and there would thus be no highest good there.[25]

It seems best to understand Kant, and other philosophers who have since written in the same vein,[26] as being concerned not with the question of what intrinsic value rational beings have – in the sense of "intrinsic value" identified above – but with the quite different question of how we ought to behave toward such creatures.

2. DOUBTS ABOUT THE CONCEPT OF INTRINSIC VALUE

In the history of philosophy, relatively few seem to have entertained doubts about the concept of intrinsic value. Much of the debate about intrinsic value has tended to be about what things actually do have such value. However, once questions about the concept itself were raised, doubts about its metaphysical implications, its moral significance, and even its very coherence began to appear.

Consider, first, the metaphysics underlying ascriptions of intrinsic value. It seems safe to say that, before the twentieth century, most moral philosophers presupposed that the intrinsic goodness of something is a genuine property of that thing, one that is no less real than the properties (of being pleasant, of satisfying a need, or whatever) in virtue of which the thing in question is good. (Several dissented from this view, however. Especially well known for their dissent are Thomas Hobbes [1588-1679], who believed the goodness or badness of something to be constituted by the desire or aversion that one may have regarding it, and David Hume [1711-1776], who similarly took all ascriptions of value to involve

---

[23] *Ibid.*, pp. 95 (Ak. 64) and 102 (Ak. 77).

[24] Kant could avoid this implication only by positing the existence of something infinitely bad whose value would counterbalance that of persons, or by denying that the world's value is proportional to that of its contents. He nowhere indicates that he is prepared to make either move.

[25] Immanuel Kant, *Lectures on Ethics* (Cambridge: Cambridge University Press, 1997), p. 44.

[26] E.g., Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, Mass.: Harvard University Press, 1993).

projections of one's own sentiments onto whatever is said to have value.[27]) It was not until Moore argued that this view implies that intrinsic goodness, as a supervening property, is a very different sort of property (one that he called "nonnatural") from those (which he called "natural") upon which it supervenes, that doubts about the view proliferated.

One of the first to raise such doubts and to press for a view quite different from the prevailing view was Axel Hägerström [1868-1939], who developed an account according to which ascriptions of value are neither true nor false.[28] This view has come to be called "noncognitivism." The particular brand of noncognitivism proposed by Hägerström is usually called "emotivism," since it holds (in a manner reminiscent of Hume) that ascriptions of value are in essence expressions of emotion. (For example, an emotivist might claim that to say "*A* is good" is not to make a statement about *A* but to say something like "Hooray for *A*!") This view was taken up by several philosophers, including most notably A. J. Ayer [1910-1989] and Charles L. Stevenson [1908-1979].[29] Other philosophers have since embraced other forms of noncognitivism. R. M. Hare [1919-2002], for example, advocated the theory of "prescriptivism" (according to which moral judgments, including judgments about goodness and badness, are not descriptive statements about the world but rather constitute a kind of command as to how we are to act),[30] and Simon Blackburn and Allan Gibbard have since proposed yet other versions of noncognitivism.[31]

Hägerström characterized his own view as a type of "value-nihilism," and many have followed suit in taking noncognitivism of all kinds to constitute a rejection of the very idea of intrinsic value. But this seems to be a mistake. We should distinguish questions about *value* from questions about *evaluation*. Questions about value fall into two main groups: *conceptual* questions (of the sorts discussed in the contributions to this volume), and *substantive* questions (concerning which see Section 6 below). Questions about evaluation have to do with what precisely is going on when *we ascribe* value to something. Cognitivists claim that our ascriptions of value constitute statements that are either true or false; noncognitivists deny this. But even noncognitivists must recognize that our ascriptions of value fall into two fundamental classes – ascriptions of intrinsic value, and ascriptions of extrinsic value – and so they too must concern themselves with the very same conceptual and substantive questions about value (What is it for something to be valuable for its own sake rather than for the sake of something else to which it is

---

[27] Thomas Hobbes, *Leviathan*, 1651 (of which there are many editions); David Hume, *A Treatise of Human Nature*, 1739 (of which there are also many editions).

[28] Axel Hägerström, *Inquiries into the Nature of Law and Morals* (Uppsala: Uppsala University Press, 1953).

[29] A. J. Ayer, *Language, Truth, and Logic* (London: Victor Gollancz, 1946); Charles L. Stevenson, *Ethics and Language* (New Haven: Yale University Press, 1944).

[30] Richard M. Hare, *The Language of Morals* (Oxford: Oxford University Press, 1952).

[31] Simon Blackburn, *Spreading the Word* (Oxford: Oxford University Press, 1984); Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, Mass.: Harvard University Press, 1990).

related? To what sort or sorts of thing can intrinsic value be sensibly ascribed? To what thing or things can intrinsic value be accurately ascribed? And so on) as cognitivists address. It may be that noncognitivism dictates or rules out certain answers to these questions that cognitivism does not, but that is of course quite a different matter from rejecting the very idea of intrinsic value on metaphysical grounds.

Another type of metaphysical challenge to intrinsic value stems from the theory of "pragmatism," especially in the form advanced by John Dewey [1859-1952].[32] The pragmatist sees the world as ever-changing; the solution to one problem becomes the source of another, what is an end in one context is a means in another, and it is therefore a mistake to seek or offer a timeless list of intrinsic goods and evils, of ends to be achieved or avoided for their own sakes. This is a theme taken up by Monroe Beardsley [1915-1985] in Chapter 7, which opens Part II of this volume. Taking his cue from Dewey, Beardsley attacks the notion of intrinsic value, arguing that all value is extrinsic, and denying that the existence of something with extrinsic value presupposes the existence of something else with intrinsic value. (In the course of his argument, Beardsley rejects the sort of "dialectical demonstration" of intrinsic value that we tried to provide in the last section when seeking an explanation of the derivative value of helping others in time of need in terms of some nonderivative value.) A quick response to Dewey's and Beardsley's misgivings about intrinsic value would be to admit that it may well be that, the world being as complex as it is, nothing is such that its value is wholly intrinsic; perhaps whatever has intrinsic value also has extrinsic value, and of course many things that have extrinsic value will have no (or, at least, neutral) intrinsic value. Far from repudiating the notion of intrinsic value, though, this admission would confirm its legitimacy. But both Dewey and Beardsley would presumably insist that this quick response misses the point of their attack, and that it really is the case, not just that whatever has value has extrinsic value, but also that nothing has intrinsic value. We leave it to you to judge whether Beardsley's argument for this claim is successful. It should be noted that, even if it is successful, the argument leaves untouched the question whether something *could* have intrinsic value. If the answer to this question is "yes," then the legitimacy of the concept of intrinsic value is once again confirmed, even if the concept has no application to anything that in fact exists.

As we have noted, some philosophers do indeed doubt the legitimacy, the very coherence, of the concept of intrinsic value. Before we turn to a discussion of this issue, however, let us for the moment presume that the concept is coherent and address a different sort of doubt: the doubt that the concept has any great moral significance. Recall the suggestion, mentioned in the last section, that discussions of intrinsic value have been compromised by a failure to distinguish certain concepts. This suggestion is at the heart of Chapter 8, written by Christine Korsgaard. She notes that intrinsic value has traditionally been contrasted with "instrumental value"

---

[32] John Dewey, *Human Nature and Conduct* (New York: H. Holt, 1922).

(the value that something has in virtue of being a means to an end) and claims that this approach is misleading. Instrumental value, she says, is to be contrasted with "final value," that is, the value that something has as an end or for its own sake; however, intrinsic value (the value that something has in itself, that is, in virtue of its intrinsic, nonrelational properties) is to be contrasted with extrinsic value (the value that something has in virtue of its extrinsic, relational properties). (An example of a nonrelational property is the property of being round; an example of a relational property is the property of being loved.) Given these two distinctions, Korsgaard maintains, it is possible that something be valuable for its own sake but not in itself; indeed, this is often actually the case. Once this fact is recognized, the moral significance of intrinsic value is put into question, since (as is apparent from our discussion so far) it is with the notion of something's being valuable for its own sake that philosophers have traditionally been, and continue to be, primarily concerned. This shift of emphasis from intrinsic to final value is also the main theme both in Chapter 9, written by Shelly Kagan, and Chapter 10, written by Wlodek Rabinowicz and Toni Rønnow-Rasmussen.

There is an important corollary to drawing a distinction between intrinsic value and final value (and between extrinsic value and nonfinal value), and that is that, contrary to what Korsgaard herself initially says, it may be a mistake to contrast final value with instrumental value. If it is possible, as Korsgaard claims, that final value sometimes supervenes on extrinsic properties, then it might be possible that it sometimes supervenes in particular on the property of being a means to some end. Indeed, Korsgaard herself suggests this when she says that "certain kinds of things, such as luxurious instruments,…are valued for their own sakes under the condition of their usefulness."[33] Kagan also tentatively endorses this idea. If the idea is coherent, then we should in principle distinguish two kinds of instrumental value, one final and the other nonfinal.[34] If something *A* is a means to something else *B* and has instrumental value in virtue of this fact, such value will be nonfinal if it is merely derivative from or reflective of *B*'s value, whereas it will be final if it is nonderivative, that is, if it is a value that *A* has in its *own* right (due to the fact that it is a means to *B*), irrespective of any value that *B* may or may not have in *its* own right.

Even if it is agreed that it is final value that is central to the concerns of moral philosophers, we should be careful in drawing the conclusion that intrinsic value is not central to their concerns. First, as Kagan notes, there is no necessity that the term "intrinsic value" be reserved for the value that something has in virtue of its intrinsic properties; presumably it has been used by many writers simply to refer to what Korsgaard calls final value, in which case the moral significance of (what is thus called) intrinsic value has of course not been thrown into doubt. Nonetheless, it should probably be conceded that "final value" is a more suitable term than

---

[33] See p. 89 below.

[34] In "Instrumental Values – Strong and Weak" (*Ethical Theory and Moral Practice*, 5 (2002): 23-43), Toni Rønnow-Rasmussen pursues the implications of this distinction in detail.

"intrinsic value" to refer to the sort of value in question, since the latter term certainly does suggest value that supervenes on intrinsic properties. But here a second point can be made, and that is that, even if use of the term "intrinsic value" is restricted accordingly, it is arguable that, contrary to Korsgaard's contention, all final value does after all supervene on intrinsic properties alone; if that were the case, then, even if there is a conceptual distinction between final and intrinsic value, and even if it is conceded that it is final value that is of central concern, still there would be little harm in continuing to use the traditional term "intrinsic value" to refer to such value. Both Kagan and Rabinowicz and Rønnow-Rasmussen explicitly discuss, and reject, the suggestion that final and intrinsic value coincide in this way; this is an issue taken up in several of the contributions to Part III of this volume.

In light of the matter just discussed, we must now decide what terminology to employ. We believe that it is clear that the main concern of moral philosophers since ancient times has been with the distinction between the value that something has for its own sake (the sort of nonderivative value that Korsgaard calls "final value") and the value that something has for the sake of something else to which it is related in some way. However, given the weight of tradition, we will continue to use the terms "intrinsic value" and "extrinsic value" to refer to these two types of value, despite Korsgaard's misgivings, and without intending thereby to endorse, or reject, the view that final value supervenes on intrinsic properties alone.

Let us now turn to doubts about the very coherence of the concept of intrinsic value, so understood. In *Principia Ethica* and elsewhere, Moore embraces the consequentialist view (mentioned above) that whether an action is morally right or wrong turns exclusively on whether its consequences are intrinsically better than those of its alternatives. Some have argued that ascribing intrinsic value to consequences in this way is fundamentally misconceived. For example, Philippa Foot maintains that talk of the goodness or badness of a state of affairs makes sense only in the context of a moral theory (of right and wrong, or of virtue and vice) that is already in place and in terms of which ascriptions of value can be understood. Goodness and badness, she contends, are therefore not free-standing concepts, contrary to the traditional presupposition that would appear to characterize the work of a great many philosophers (not just consequentialists) since ancient times.[35] In Chapter 11 Judith Thomson takes up this thesis and elaborates on it considerably. Setting out from Peter Geach's idea (echoed in Foot's paper) that nothing is just plain good or bad,[36] Thomson argues that all goodness is "goodness in a way," and that this shows that appeals to intrinsic value (of the sort made by Moore but also by many nonconsequentialists) are conceptually confused. In Chapter 12 Michael Zimmerman rebuts Thomson's attack, arguing that it misidentifies its target and thus misses its mark.

---

[35] Philippa Foot, "Utilitarianism and the Virtues." *Mind*, 94 (1985): 196-209.
[36] Peter Geach, "Good and Evil." *Analysis*, 17 (1956): 33-42.

3. THE BEARERS OF INTRINSIC VALUE

Among those who do not doubt the coherence of the concept of intrinsic value there is considerable difference of opinion about what sort or sorts of entity can have such value. Moore does not explicitly address this issue, but his writings show him to have a liberal view on the matter. There are times when he talks of individual objects (e.g., books) as having intrinsic value,[37] others when he talks of the consciousness of individual objects (or of their qualities) as having intrinsic value,[38] others when he talks of the existence of individual objects as having intrinsic value,[39] others when he talks of types of individual objects as having intrinsic value,[40] and still others when he talks of states of individual objects as having intrinsic value.[41]

Moore would thus appear to be a "pluralist" concerning the bearers of intrinsic value. Others take a more conservative, "monistic" approach, according to which there is just one kind of value bearer. Ross, for example, maintains that it is at bottom what he calls objectives or facts, and only such entities, that can have intrinsic value.[42] Facts are the sort of thing to which we refer by certain "that"-clauses, such as "It is good that John is pleased" (or "That John is pleased is good"). According to Ross, the apparent attribution of intrinsic value to things other than facts can be reduced to, or "translated into," the attribution of such value to facts. For example, the claim that a particular rare stamp has intrinsic value might be recast as the claim that it is intrinsically good that the stamp is rare.[43] Whether such reductions are acceptable has been a matter of considerable debate. Proponents maintain that it introduces some much-needed order into the discussion of intrinsic value; opponents charge that it results in distortion and oversimplification. The papers in Part III of this volume all deal with this issue.

In Chapter 13 Chisholm adopts and elaborates on the view (espoused by Ross but embraced even earlier by Alexius Meinong [1853-1920], whose ideas Chisholm discusses at some length) that it is objectives that are the bearers of intrinsic value. (In the course of his discussion, Chisholm also touches on many related issues, having to do with the logic and the computation of intrinsic value, which are the focus of his contributions to Parts IV and V of this volume and which we will address in the next two sections.) In Chapter 14 Lemos likewise adopts, at least tentatively, the monistic view that facts are the only things that can have intrinsic value. However, in so doing, he introduces a subtle twist to Chisholm's position. Whereas Chisholm takes objectives to be abstract states of affairs (such as the state of affairs of everyone being happy) that may or may not obtain, Lemos insists that

---

[37] Moore, *Principia Ethica*, p. 3.
[38] *Ibid.*, p. 195.
[39] *Ibid.*, p. 196.
[40] Moore, *Philosophical Studies*, p. 260.
[41] Moore, *Principia Ethica*, p. 195.
[42] W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press), pp. 112-13.
[43] The example is borrowed from Monroe Beardsley, "Intrinsic Value," pp. 61-62 below.

the bearers of intrinsic value are states of affairs that *do* obtain. His reason for doing so is that it is only in virtue of a state of affairs' obtaining that there is anything of value in the world. (For example, he would say that there is nothing good about the state of affairs of everyone being happy, since unfortunately it does not obtain.) Like Chisholm, though, Lemos takes the bearers of value to be abstract rather than concrete.

This raises a difficult issue: how to distinguish abstract from concrete entities. There is no consensus on the matter. Although everyone agrees that individual objects, such as tables and chairs, are concrete and that the sort of states of affairs that Chisholm discusses are abstract, how to characterize in these terms other things such as actual events, states, processes, lives, or the sort of facts that Lemos discusses, is controversial. In Chapter 15 Zimmerman remains silent on this issue while arguing for the monistic view that it is not individual objects, but only states of individual objects, that can have intrinsic value. He contends, furthermore, that on this view the value that something has for its own sake will always supervene on its intrinsic properties alone, so that the traditional term "intrinsic value" is perfectly appropriate in this context. In Chapter 16 Torbjörn Tännsjö also argues for a monistic view, according to which it is only states or processes that can have intrinsic value; but he is adamant that these entities are to be understood as concrete, claiming that this allows for a straightforward application of Moore's isolation test to the question of what things have value and in what degrees they have it.

In contrast to the monism that characterizes the four chapters just discussed, Rabinowicz and Rønnow-Rasmussen advocate a pluralistic approach in Chapter 10, their first contribution to this volume. There they discuss, and reject, a number of attempts to reduce ascriptions of value to various types of entities to ascriptions to just one type of entity. In Chapter 17, their second contribution, they return to this issue, addressing a new reductionist proposal according to which the ultimate bearers of value are "tropes" (particularized properties such as the rectangularity and the whiteness of this page). They claim that this proposal has greater merit than the ones they addressed in their earlier paper but that, in the final analysis, it too is to be rejected. They thus remain committed to pluralism about the bearers of intrinsic value.

4. THE LOGIC OF INTRINSIC VALUE

Whether we are monists or pluralists regarding the bearers of intrinsic value, we will surely all agree that something that is intrinsically good is intrinsically better than something that is intrinsically neutral, which in turn is intrinsically better than something that is intrinsically bad. Comparisons of intrinsic value often require that we reach beyond this platitude, however, and it is here that a thorough understanding of the logic of intrinsic value can be especially useful.

It was not until the twentieth century that philosophers began to investigate this logic in earnest. At that time great strides were being made in the study of logic generally. Gottlob Frege [1848-1925], often called the founder of modern logic,

broke with the Aristotelian tradition that still prevailed in his day when he introduced the quantifier-variable notation for expressing generality. This has fueled the work of logicians ever since, driving the development of both propositional and predicate logic as well as giving rise to some more specialized offshoots, such as the logics of necessity and possibility ("modal logic"), of obligation and permission ("deontic logic"), of past, present, and future ("temporal logic"), of knowledge and belief ("epistemic logic"), and others – including the logic of intrinsic value.

Important work on the logic of intrinsic value was accomplished by several philosophers, including perhaps most notably Sören Halldén and Georg Henrik von Wright [1916-2003].[44] However, a turning point came with the publication in 1966 of a ground-breaking paper by Chisholm and Ernest Sosa.[45] In this paper, the authors succinctly demonstrated that a number of principles that featured in the accounts of previous writers were in fact unacceptable. (Among these principles are the following: that a state of affairs is intrinsically good if and only if it is intrinsically better than its negation, and that one state of affairs is better than another if the negation of the latter is intrinsically better than the negation of the former.) In order to improve on these accounts, Chisholm and Sosa presented a rigorous treatment of the logic of intrinsic value that built on standard propositional logic. Taking the notion of intrinsic betterness as primitive, they provided definitions of certain key concepts (sameness in intrinsic value, intrinsic indifference, intrinsic neutrality, intrinsic goodness, and intrinsic badness), supplemented these definitions with five axioms concerning intrinsic betterness, stipulated certain rules of inference, and then derived forty-three theorems about intrinsic value, each of which they claimed to be intuitively plausible. Their work provoked a number of responses. Chisholm himself sought in later papers to improve upon and extend the account that he and Sosa had given. Two such papers are Chapters 1 and 13, his contributions to Parts I and III of this volume, respectively. Another such paper is Chapter 18, his contribution to Part IV. In all these papers, Chisholm provides a clear summary of the main points made in the original paper written by himself and Sosa,[46] and so there is no need to reproduce these points here. In Chapter 18 Chisholm tackles an issue that he and Sosa did not address but which has preoccupied philosophers considerably since, that of the intrinsic value to be attributed to disjunctive states of affairs. An example of such a state of affairs is its being the case that *either* Jones is pleased *or* Smith is displeased. If we assume that it is (or would be) intrinsically good that Jones is pleased and intrinsically bad that Smith is displeased, what value should we attribute to the disjunction of these two states? Neutrality? What if it should happen that the extent of pleasure that is at stake is greater than the extent of displeasure? Should that incline us to think that the disjunction is intrinsically good? It is with such questions that Chisholm's paper deals.

---

[44] See Sören Halldén, *On the Logic of 'Better'* (Lund and Copenhagen: Library of *Theoria* 2, 1957); Georg Henrik von Wright, *The Logic of Preference* (Edinburgh: Edinburgh University Press, 1963).
[45] Roderick Chisholm and Ernest Sosa, "On the Logic of 'Intrinsically Better'." *American Philosophical Quarterly*, 3 (1966): 244-49.
[46] See pp. 6-7, 174-75, and 232 below.

In Chapter 19 Philip Quinn [1940-2004] proposes an adjustment to Chisholm's account. Chisholm assumes that all states of affairs are comparable in intrinsic value, that is, that, for any two states, either one is intrinsically better than the other or both have the same intrinsic value. Quinn claims that this is to assume too much, since it is at least arguable that some states of affairs are incomparable in intrinsic value. (He asks us to contrast two comparisons. Consider comparing the enjoyment of the taste of apples with that of the taste of pears. Now consider comparing the former with the enjoyment of the sound of Beethoven's Ninth Symphony. Whereas the first comparison seems straightforward, the latter does not and might be thought to involve two items that are in fact strictly incomparable.) Quinn proposes an account of intrinsic value that is akin to Chisholm's but whose axioms do not assume comparability. Quinn's account does, however, have comparability as a theorem; that is, given his axioms, it can be proven (rather than simply assumed) that no two states of affairs are incomparable in intrinsic value. Whether this is an acceptable result is controversial. We will see in the next section that some philosophers reject it.

Almost all writers on intrinsic value (including both Chisholm and Quinn) presuppose that the relation of intrinsic betterness is transitive, that is, that if one state is intrinsically better than another which is itself intrinsically better than a third, then the first is intrinsically better than the third. This very natural assumption has recently been called into question. In Chapter 20 Stuart Rachels argues that there are cases in which one state is better than a second which is itself better than a third, but in which the first is not better than the third. He puts his thesis in terms of "all things considered better than," but he would also apply it to "intrinsically better than."[47] In Chapter 21 Ken Binmore and Alex Voorhoeve maintain that Rachels's argument fails for the same sort of reason that Zeno's argument that Achilles could not overtake the tortoise fails. However, in Chapter 22 Erik Carlson points up a limitation to this criticism of Rachels, thus leaving the door open to opponents of transitivity.

5. THE COMPUTATION OF INTRINSIC VALUE

In our assessments of intrinsic value, we are often and understandably concerned not only with *whether* something is good or bad but with *how* good or bad it is. Arriving at an answer to the latter question is not straightforward. At least three problems threaten to undermine the computation of intrinsic value.

First, there is the possibility, just mentioned, that the relation of intrinsic betterness is not transitive. Should this be the case, it would seriously complicate comparisons, and hence assessments, of intrinsic value.

Second, there is the possibility (raised by Quinn in Chapter 19) that certain values are incommensurate. Ross, for example, has this to say on the subject:

[47] See Stuart Rachels, "Intransitivity," in *Encyclopedia of Ethics*, second edition, edited by L. C. Becker and C. B. Becker (New York: Routledge, 2001), pp. 877-79.

> [W]e are quite incapable of equating, in respect of goodness, any amount of pleasure with any amount of morally good action. I suggested in *The Right and the Good* that while both virtue and pleasure have places on the same scale of goodness, virtue begins at a higher point than that at which pleasure leaves off, so that any, even the smallest, amount of virtue is better, and more worth bringing into existence, than any, even the greatest, amount of pleasure. But I now see this...to be impossible. If virtue really were on the same scale of goodness as pleasure, then pleasure of a certain intensity, if enjoyed by a sufficiently large number of persons or for a sufficient time, would counterbalance virtue possessed or manifested only by a small number or only for a short time. But I find myself quite unable to think this to be the case; and if I am right in this, it follows that pleasures, if ever good, must be good in a different sense from that in which good activities are so.[48]

There is some confusion here. In claiming that virtue and pleasure are incommensurate for the reason given, Ross presumably means that they cannot be measured on the same *ratio* scale. (A ratio scale is one with an arbitrary unit but a fixed zero point. Mass and length are standardly measured on ratio scales.) But incommensurability on a ratio scale does not imply incommensurability on *every* scale – an ordinal scale, for instance. (An ordinal scale is simply one that supplies an ordering for the quantity in question, such as the measurement of arm-strength that is provided by an arm-wrestling competition.) Ross's remarks indicate that he in fact believes that virtue and pleasure *are* commensurate on an ordinal scale, since he appears to subscribe to the arch-puritanical view that any amount of virtue is intrinsically better than any amount of pleasure. This view is just one example of the thesis that some goods are "higher" than others, in the sense that the former "begin" (as Ross puts it) at a "higher point" than that at which the latter "leave off." This thesis can be traced to the ancient Greeks,[49] and it has been endorsed by many philosophers since, perhaps most famously by John Stuart Mill [1806-1873].[50] Interest in the thesis has recently been revived by a set of intricate and intriguing puzzles, posed by Derek Parfit, concerning the relative values of low-quantity/high-quality goods and high-quantity/low-quality goods.[51] One response to these puzzles is to adopt Rachels's thesis of nontransitivity. (Parfit himself eschews this response.[52]) Another response is to insist on the thesis that some goods are higher than others. (This is not to say that subscription to this thesis by itself solves the puzzles that Parfit raises.) In Chapter 23 Neil Feit proposes a way of reconciling this thesis about higher goods with a summative approach to the computation of intrinsic value, an approach with which it might at first seem to be at odds. In Chapter 24 Gustaf Arrhenius argues that the thesis has the surprising implication that, if there are two types of goods *A* and *B* such that any amount of *A* is better than any amount

[48] Ross, *Foundations of Ethics*, p. 275.
[49] See Plato, *Philebus*, 21a-e; Aristotle, *op. cit.*, 1174a.
[50] See paras. 4 ff. of his *Utilitarianism*, 1863 (of which there are many editions).
[51] Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), Part IV.
[52] *Ibid.*, p. 441.

of *B*, then there must be two types of goods *C* and *D* such that some amount of *C* is better than any amount of *D*, even though goods of type *C* are only marginally better than goods of type *D*.

To repeat: contrary to what Ross says, the thesis that some goods are higher than others implies that such goods are commensurate, and not that they are incommensurate. Some people do hold, however, that certain values really are incommensurate, in that they cannot be compared on any meaningful scale. (Isaiah Berlin [1909-1997], for example, is often thought to have said this about the values of liberty and equality. Whether he is best interpreted in this way is debatable.[53]) This view constitutes a more radical threat to the computation of intrinsic value than does the view that intrinsic betterness is not transitive. The latter view presupposes at least some measure of commensurability. If *A* is better than *B* and *B* is better than *C*, then *A* is commensurate with *B* and *B* is commensurate with *C*; and even if it should turn out that *A* is not better than *C*, it may still be that *A* is commensurate with *C* (either because it is as good as *C* or because it is worse than *C*). But if *A* is incommensurate with *B*, then *A* is neither better than nor as good as nor worse than *B*. If such a case can arise, there is an obvious limit to the extent to which we can meaningfully say how good a certain complex whole is (here, "whole" is used to refer to whatever kind of entity may have intrinsic value); for, if such a whole comprises incommensurate goods *A* and *B*, then there will be no way of establishing how good it is overall, even if there is a way of establishing how good it is with respect to each of *A* and *B*.

There is a third, still more radical threat to the computation of intrinsic value. Quite apart from any concern with the commensurability of values, Moore famously claims that there is no easy formula for the determination of the intrinsic value of complex wholes because of the truth of what he calls the "principle of organic unities."[54] According to this principle, the intrinsic value of a whole must not be assumed to be the same as the sum of the intrinsic values of its parts.[55] As an example of an organic unity, Moore gives the case of the consciousness of a beautiful object; he says that this has great intrinsic value, even though the consciousness as such and the beautiful object each have comparatively little, if any, intrinsic value. If the principle of organic unities is true, then there is scant hope of a systematic approach to the computation of intrinsic value. Although the principle explicitly rules out only summation as a method of computation, Moore's remarks strongly suggest that there is no relation between the parts of a whole and the whole itself that holds in general and in terms of which the value of the latter can be computed by aggregating (whether by summation or by some other means) the values of the former. Moore's position has been endorsed by many other philosophers. For example, Ross says that it is better that one person be good and happy and another bad and unhappy than that the former be good and unhappy and

---

[53] Isaiah Berlin, *Four Essays on Liberty* (Oxford: Oxford University Press, 1969).
[54] Moore, *Principia Ethica*, p. 96.
[55] *Ibid.*, p. 28.

the latter bad and happy, and he takes this to be confirmation of Moore's principle.[56] Broad takes organic unities of the sort that Moore discusses to be just one instance of a more general phenomenon that he believes to be at work in many other situations, as when, for example, two tunes, each pleasing in its own right, make for a cacophonous combination.[57] Others have furnished still further examples of organic unities.

Was Moore the first to call attention to the phenomenon of organic unities in the context of intrinsic value? This is debatable. Despite the fact that he explicitly invoked what he called a "principle of summation" that would appear to be inconsistent with the principle of organic unities,[58] Brentano appears nonetheless to have anticipated Moore's principle in his discussion of *Schadenfreude*, that is, of malicious pleasure; he condemns such an attitude, even though he claims that pleasure as such is intrinsically good.[59] Certainly Chisholm takes Brentano to be an advocate of organic unities. In Chapter 25 Chisholm ascribes to Brentano the view that there are many kinds of organic unity and builds on what he takes to be Brentano's insights (and, going further back in the history of philosophy, the insights of St. Thomas Aquinas [1225-1274] and Arthur Schopenhauer [1788-1860]). He ends by providing a list of ten different ways in which an organic unity may be constituted. In Chapter 26 Lemos finds fault with some of the details of Chisholm's view, although he is very much taken with Chisholm's general approach and is himself an advocate of organic unities.[60]

Another fan of the principle of organic unities is Dancy, who puts a special spin on it. In Chapter 27 he outlines a radical approach to the assessment of value according to which the intrinsic value of something may vary from context to context; indeed, the variation may be so great that the thing's value changes "polarity" from good to bad, or *vice versa*. In keeping with Korsgaard and others mentioned in Section 2 above, Dancy holds that something's intrinsic value need not supervene on its intrinsic properties alone; in fact, the supervenience-base may be so open-ended that it resists generalization. This "particularist" approach to value (an approach that Dancy also applies to reasons) constitutes an endorsement of the principle of organic unities that is even more subversive of the computation of intrinsic value than Moore's. Moore holds that the intrinsic value of something is and must be constant, even if its contribution to the value of wholes of which it forms a part is not. Dancy holds that something's intrinsic value may itself be inconstant.

Not everyone has accepted the principle of organic unities; some have held out hope for a more systematic approach to the computation of intrinsic value. However, even someone who is inclined to measure intrinsic value in terms of summation must acknowledge that there is a sense in which the principle of organic unities is

---

[56] Ross, *The Right and the Good*, p. 72.
[57] C. D. Broad, *Ethics* (Dordrecht: Martinus Nijhoff, 1985), p. 256.
[58] Brentano, *op. cit.*, p. 23n.
[59] *Ibid.*, p. 90.
[60] Lemos, *op. cit.*, ch. 3.

obviously true. Consider some complex whole, *W*, that is composed of three goods, *X*, *Y*, and *Z*, that are wholly independent of one another. Suppose that we had a ratio scale on which to measure these goods, and that their values on this scale were 10, 20, and 30, respectively. We would expect someone who takes intrinsic value to be summative to declare the value of *W* to be (10 + 20 + 30 =) 60. But notice that, if *X*, *Y*, and *Z* are parts of *W*, then so too, presumably, are the combinations *X*-and-*Y*, *X*-and-*Z*, and *Y*-and-*Z*; the values of these combinations, computed in terms of summation, will be 30, 40, and 50, respectively. If the values of these parts of *W* were also taken into consideration when evaluating *W*, the value of *W* would balloon to 180. Clearly, this would be a distortion. Someone who wishes to maintain that intrinsic value is summative must thus show not only how the various alleged examples of organic unities provided by Moore and others are to be reinterpreted, but also how, in the sort of case just sketched, it is only the values of *X*, *Y*, and *Z*, and not the values either of any combinations of these components or of any parts of these components, that are to be taken into account when evaluating *W* itself. In order to bring some semblance of manageability to the computation of intrinsic value, this is precisely what some writers, by appealing to the idea of "basic" intrinsic value, have tried to do. The general idea is this. In the sort of example just given, each of *X*, *Y*, and *Z* is to be construed as having basic intrinsic value; if any combinations or parts of *X*, *Y*, and *Z* have intrinsic value, this value is not basic; and the value of *W* is to be computed by appealing only to those parts of *W* that have basic intrinsic value.

Gilbert Harman was one of the first explicitly to discuss basic intrinsic value. In Chapter 28 he points out the apparent need to invoke such value if we are to avoid distortions in our evaluations. However, he offers no precise account of the concept of basic intrinsic value and ends his paper by saying that he can think of no way to confirm or disconfirm that it would suffice to put our computational house in order. In Chapter 29 Carlson tentatively attempts to put this concern to rest. On the assumption that an acceptable account of basic intrinsic value can be given, Carlson discusses how the computation of nonbasic intrinsic value might be accomplished. He criticizes proposals on this matter made by Warren Quinn [1938-1991] and Edward Oldfield and then offers three proposals of his own, the last of which builds on the idea that there can be states that have no determinate intrinsic value but which are nonetheless comparable in value with other states. In Chapter 30 Sven Danielsson discusses not only Harman's paper but also those of Quinn, Oldfield, and Carlson. He finds reason to disagree on some points with all four of these authors and offers his own proposal on how to compute nonbasic intrinsic value.

None of the contributions so far discussed offers a detailed characterization of basic intrinsic value. This is what Feldman attempts to provide in Chapter 31. Claiming that invoking such value is useful for solving a number of puzzles about not only the computation of intrinsic value but also certain related matters, Feldman identifies six main features that any state with basic intrinsic value must possess and then goes on to propose solutions to the puzzles in question. At no point, however, does he offer an analysis of the concept of basic intrinsic value. In Chapter 32

Zimmerman proposes an analysis of this concept. By distinguishing between what he calls "actual intrinsic value" and "virtual intrinsic value," Zimmerman attempts in one fell swoop not only to give the analysis in question but also to show why the principle of organic unities is to be rejected (once it has been properly formulated) and, furthermore, to show how some of the issues about the computation of intrinsic value (such as how to compute the value of disjunctive states) which have exercised many philosophers are in fact mere pseudo-problems that dissolve upon inspection.

A final note to this section: we are now in a position to explain why we said in Section 1 that perhaps not all intrinsic value is nonderivative. If it is correct to distinguish between basic and nonbasic intrinsic value and also to compute the latter in terms of the former, then there is clearly a respectable sense in which nonbasic intrinsic value is derivative.

6. INSTANCES OF INTRINSIC VALUE

In Section 2 above, we distinguished two categories of questions about intrinsic value: *conceptual* questions (of the sort discussed in the contributions to this volume) and *substantive* questions, that is, questions about what things may be accurately said to have intrinsic value and to what degree. There is no doubt that it is with the substantive questions that philosophers have traditionally been primarily concerned; as we mentioned at the outset, it is only within the last one hundred years or so that the underlying conceptual questions have been intensively investigated. There is a sense in which the conceptual questions come first; if they are not answered, any answers to the substantive questions must be regarded as at best tentative. Nonetheless, there is also a clear sense in which the substantive questions are more pressing, since it is the answers to them that will have a direct implication regarding how we are to evaluate our circumstances and live our lives.

A great deal has been written on the substantive questions – far too much to be represented in this volume alongside the contributions that we have selected. In this introduction, we have already given you a glimpse of what some philosophers have found to be intrinsically valuable: pleasure, knowledge, love, and the like. There is considerable disagreement about just what should be featured on our list of intrinsic goods. One of the most comprehensive lists that anyone has suggested is that proposed by William Frankena [1908-1994]. It is this: life, consciousness, and activity; health and strength; pleasures and satisfactions of all or certain kinds; happiness, beatitude, contentment, etc.; truth; knowledge and true opinions of various kinds, understanding, wisdom; beauty, harmony, proportion in objects contemplated; aesthetic experience; morally good dispositions or virtues; mutual affection, love, friendship, cooperation; just distribution of goods and evils; harmony and proportion in one's own life; power and experiences of achievement; self-expression; freedom; peace, security; adventure and novelty; and good

reputation, honor, esteem, etc.[61] (Presumably a corresponding list of intrinsic evils could be provided.) Regarding this inventory of intrinsic goods, Frankena says:

> Religious experiences or values, which many rate highest among intrinsic goods, are not mentioned separately in this list because they presuppose the existence of God and so raise questions that cannot be dealt with here. The communion with and love or knowledge of God that Augustine and Aquinas regard as the highest good would, however, presumably come under our headings of knowledge and love. Other intrinsically good religious experiences would also probably fall under these or other headings.[62]

He adds:

> Reflecting on the...list of proposed intrinsic values myself, I come to the following conclusions. It seems to me that all of them may be kept on the list, and perhaps others may be added, if it is understood that it is the *experience* of them that is good in itself.[63]

Comprehensive as it is, there nonetheless appears to be one conspicuous omission from Frankena's list, namely, the increasingly popular view that certain environmental entities or qualities have intrinsic value. Some find intrinsic value, for example, in certain "natural" environments (wildernesses untouched by human hand); some find it in certain animal species; and so on. (It is of course possible that Frankena would say that, like the love and knowledge of God, these matters are implicitly represented by one or more items that are already on the list.)

For this and other reasons the list is controversial, as is Frankena's qualification that it is the experience of the items listed that has intrinsic value. Many philosophers would amend the list in some way; many would deny that experience of the items in question is required for there to be something of intrinsic value. In one way or another, though, almost all the assertions that have been made by philosophers through the ages that such-and-such has intrinsic value are reflected somewhere on Frankena's list. (Among these philosophers are such highly influential figures as Plato, Aristotle, Epicurus, St. Augustine [354-430], St. Thomas Aquinas, Thomas Hobbes, John Locke [1632-1704], Joseph Butler [1692-1752], Bentham, Mill, Henry Sidgwick [1838-1900], Brentano, Moore, Ross, and Broad, to name only a few.) We hope that attention to the conceptual questions addressed in this volume will help pave the way to a resolution of the fundamental substantive question concerning just how, if at all, Frankena's list is to be revised.[64]

---

[61] William K. Frankena, *Ethics* (Englewood Cliffs: Prentice Hall, 1973), pp. 87-88.
[62] *Ibid*., p. 88.
[63] *Ibid*., p. 89.
[64] Thanks to Fred Feldman, Noah Lemos, Terry McConnell, and an anonymous referee for comments on earlier drafts.

NOTE TO READERS

We have altered the original contents of the works collected here in only minor ways, if at all. We have not aimed for consistency in punctuation, spelling, or the manner in which reference is made to other works; conventions vary widely, and we have judged it best to accept that they do. However, we have corrected obvious typographical errors; moreover, we have endeavored to update or otherwise fill in bibliographical references that might otherwise have been misleading. (This has sometimes had the result of an earlier work's making full reference to a later work, as if the author had rather formidable predictive powers. We trust that you won't be misled by this.) Also, on rare occasions, we have inserted editorial notes (separated from the rest of the text by square brackets and marked by an asterisk) for purposes of clarification. Finally, in keeping with the theme of this anthology, the bibliography provided at the end is restricted to works that have appeared within the last fifty years, with a pronounced emphasis on works that have appeared most recently.

PART I

IDENTIFYING THE CONCEPT OF INTRINSIC VALUE

CHAPTER 1

R. M. CHISHOLM

INTRINSIC VALUE[1]

1. INTRODUCTION

It was obvious to Plato and Aristotle and, in more recent times, to Brentano and Moore, that there is a distinction between 'intrinsic' and 'nonintrinsic' value. These philosophers took it for granted that, if there is anything that is good, then there is something that is intrinsically good or good in itself, and that if there is anything that is bad, then there is something that is intrinsically bad or bad in itself. But at the present time this distinction is often called into question and even ridiculed. In this paper I will defend the distinction.

2. THE PROBLEM OF DEFINING INTRINSIC VALUE

Our problem concerns, not the nature of value or of good and evil generally, but the nature of *intrinsic* value and of *intrinsic* goodness and *intrinsic* evil. We are concerned with the qualification 'intrinsic'. And so we might put our question this way: given the generic concept of *value,* can we define *intrinsic* value?

In raising this question, I am assuming, then, that we are given a generic sense of such value expressions as 'good,' 'bad,' 'neutral,' and 'better.' (But in setting forth the answer, I will take 'better' as the only undefined value expression.) And I also assume that states of affairs (or propositions) are bearers of value and that possible worlds are states of affairs and hence also bearers of value.

Perhaps it may be agreed, preanalytically, that that state of affairs which is *someone experiencing pleasure* is good and that, on a certain occasion, that state of affairs which is *someone undergoing discomfort* is also good. But the former state of affairs and not the latter is intrinsically good. Speaking somewhat imprecisely, we might put this point by saying that the two states of affairs differ in the following respect: the former, unlike the latter, would be good no matter what else happened. The goodness of the latter is 'extrinsic' in that it is dependent, somehow,

[1] I wish to express my indebtedness to Fred Feldman and Lars Bergström.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 1-10.
© 2005 *Springer. Printed in the Netherlands.*

upon the occurrence of something else. Aristotle and Moore both suggest that the intrinsic value of a state of affairs is the value that that state of affairs 'would have in isolation.'

But how are we to understand the 'would have' and in what sense can a state of affairs be 'isolated'?

We may be tempted to say this: a state of affairs is *intrinsically* good provided only it is good in every possible world in which it obtains, and analogously for intrinsic neutrality and intrinsic evil.[2] But this appeal to modal concepts is not sufficient to enable us to make the distinction we are looking for.

Consider, for example, a world in which some intrinsic good – say, someone's innocent pleasure – leads to an enormous amount of subsequent evil. Using 'good' in its generic sense, can we say that in such a world the pleasurable experience is good? It would hardly seem so. Hence we cannot say that an intrinsically good state of affairs is one which is good in every possible world in which it obtains. Similar considerations hold of intrinsic evil.

And it follows, of course, that we cannot use this method to characterize the relation of being *intrinsically better.* That is to say, it will not do to define '*p* is intrinsically better than *q*' by saying '*p* is better than *q* in every world in which they both obtain.' Thus we may wish to say that *Jones feeling pleasure* (*p*) is intrinsically better than *Jones undergoing discomfort* (*q*). But we cannot say, in the more generic sense of 'better,' that *p* is better than *q* in every possible world in which both *p* and *q* occur; for we may want to say, of those worlds in which *q* is productive of enormous amounts of good and *p* of enormous amounts of evil, that in those worlds *q* is better than *p*. Hence we cannot define intrinsic value merely by reference to what is valuable in every possible world.

Yet the point of saying that *Jones undergoing discomfort* is only 'extrinsically' good would seem to be: it wouldn't be good unless some *other* good state of affairs obtains which doesn't include it and isn't included in it. And the point of saying that *Jones feeling pleasure* is 'intrinsically' good would seem to be: its goodness doesn't require that there obtain some *other* good state of affairs which neither includes it nor is included within it; *Jones feeling pleasure* would be good even if, of all the states of affairs that obtain, it is included in every one that is good. And analogously for 'extrinsically' and 'intrinsically' bad.

Can we put the distinction more precisely?

3. INTRINSIC VALUE STATES

We first introduce the following intentional sense of entailment:

---

[2] I suggested this definition in 'Objectives and Intrinsic Value,' in Rudolf Haller, ed., *Jenseits von Sein und Nichtsein* (Graz: Akademisches Druck- und Verlagsanstalt, 1972), 261-9 [* pp. 171-79 of this volume]: see p. 262 [* p. 172 of this volume].

> *p* entails *q* = *Df p* is necessarily such that (i) if it obtains then *q* obtains and (ii) whoever accepts it accepts *q*.

Now we may say: a state of affairs *p* is 'part' of a state of affairs *q*, if and only if *q* entails *p*.

I believe that the intrinsic value concepts that we are looking for may be found if we consider what it is for one state of affairs to 'reflect all the good and evil' that there is in another state of affairs. I suggest that this concept may be explicated as follows, in terms of the generic concepts that we are presupposing:

> *p* reflects all the good and evil that there is in *q* = *Df q* entails *p*; and every part of *q* that entails *p* has the same value as *p*.

If *p* thus reflects all the good and evil that there is in *q*, then, since every state of affairs is a 'part' of itself, *p* will have the same value as *q*. And therefore, if there is another state of affairs *r*, such that *p* also reflects all the good and evil that there is in *r*, then *q* will have the same value as *r*. More generally, if *p* reflects all the good and evil that there is in *q*, then *p* and *q* will enter into the same preferability relations: therefore, whatever is the same in value as the one is the same in value as the other; whatever is better than the one is better than the other; and whatever is worse than the one is worse than the other.

What would be a case of a *p* and a *q* such that *p* reflects all the good and evil that there is in *q*? In order to have an illustration let us assume, for simplicity, that hedonism is true. Now we may consider that state of affairs which is *Jones feeling pleasure.* But following the suggestions of Aristotle and Moore, let us attempt to consider this state 'in isolation.'

We are viewing possible worlds as being themselves states of affairs. Hence one relevant way of considering *Jones feeling pleasure* 'in isolation' is to conceive a possible world in which Jones feels pleasure and in which there is no other pleasure or displeasure. Then, if hedonism is true, *Jones feeling pleasure* will reflect all the good and evil there is in that world; so, too, for any wider state of affairs in that world that entails *Jones feeling pleasure.* And, given the assumption of hedonism, we may say that any such world is good.[3] Analogously, we may conceive a possible world which is such that *Jones feeling displeasure* reflects all the good and evil that there is in that world. And if hedonism is true, every such world is bad.

Let us use the expression 'intrinsic value state' to refer to states of affairs such as those we have just illustrated – states of affairs which are such that, for some possible world, they reflect all the good and evil that there is in that world.

---

[3] Suppose that, for any amount of pleasure that one may experience, it is also possible to experience a lesser but still positive amount of pleasure. Then *Jones feeling pleasure* would not itself be an intrinsic value state, since any world containing it would also contain the good that is in some wider state of affairs (e.g., *Jones feeling pleasure to degree 10*). In such a case the latter state of affairs, but not the former, will be an intrinsic value state. This possibility was pointed out to me by W. Rabinowicz.

> *p* is an intrinsic value state = *Df* There is a possible world *W* such that: *p* reflects all the good and evil that there is in *W*; and if *p* is not neutral, then everything that reflects all the good and evil that there is in *W* either entails or is entailed by *p*.

Why the second clause? How could it possibly be the case that (a) *p* reflects all the good and evil that there is in *W* and yet *not* be the case that (b) everything that reflects all this good and evil entails or is entailed by *p*? It should be kept in mind that we are presupposing only a generic value concept and that we are not yet in a position to distinguish intrinsic values from instrumental values. Using this more generic value concept, we may want to say that, on occasion, the means to a valuable result may have the same generic value as the result. For example, if the patient being well is generically good, then so, too, is the treatment that brought that state of affairs about. If these two things are thought of as being the same in generic value, then the treatment as well as the patient being well would satisfy the first condition of our definiens: in a world in which they obtained together, they could *both* reflect all the good and evil that there is in that world.[4] But the treatment is such that, in order to have this property, it must be followed by a good result. That is to say, the world must be such that, in addition to the treatment, there is still *another* state of affairs – in this case, the patient feeling well – which may be said to contain all the good and evil there is in that world. But, on the other hand, that intrinsic good which is the patient feeling well *could* be one that either entails or is entailed by everything that reflects all the good and evil that there is in that world. Intrinsic value states, then, are states of affairs that may be good or bad 'in isolation.' For our definition tells us, in effect, that an intrinsic value state is capable of containing all the good and evil that there is in a world.

Many states of affairs, we should note, are not thus intrinsic value states. For example:

> Jones is pleased or Smith is displeased.
>
> Robinson is pleased and either Jones is pleased or Smith is displeased.

The first of these states of affairs is incapable of reflecting all the good and evil that there is in any possible world; for any possible world in which this disjunctive state of affairs occurs will be a world *also* containing the good and evil that is in one or the other or both of the two disjuncts. And analogously for the second state of affairs.[5]

---

[4] This was pointed out to me by Eva Bodanszky. Were it not for this possibility, we could replace '*p* reflects all the good and evil that there is in *q*' by '*p* contains all the good and evil that there is in *q*.'

[5] We could now say that intrinsic value states are the 'source' of the good and evil that is to be found in any possible world. The 'ultimate source' of such good and evil could then be said to be certain *basic* intrinsic value states – those intrinsic value states which are either good or bad and which have no intrinsic value states as proper parts. It should be noted that 'proper part' is here understood in terms of a strict sense of entailment – a sense of entailment which does not enable us to say that for any states of

Now let us distinguish two different types of intrinsic value states.

First, there are those states of affairs *p* which are such that both *p* and its negation are intrinsic value states. An example would be: Jones being pleased to degree 10.

And secondly, there are those states of affairs *p* which are such that *p* is an intrinsic value state and its negation is *not* an intrinsic value state. An example would be:

> It being false that Jones is pleased to degree 10 and it also being false that Jones is displeased to degree 10.

The negation of this state of affairs is incapable of containing all the good and evil there is in any possible world. For this negation is equivalent to the following disjunction:

> Either Jones being pleased to degree 10 or Jones being displeased to degree 10.

Every world in which this disjunction obtains will include, *in addition*, either the good that is Jones being pleased or the evil that is Jones being displeased.

Let us now restrict the *bearers of intrinsic value* to those intrinsic value states of the second sort – those which are such that they have intrinsic value states as their negations:

> *p* is a bearer of intrinsic value = *Df p* is an intrinsic value state and not-*p* is an intrinsic value state.

Now we are in a position to characterize intrinsic value.

4. DEFINITION OF INTRINSIC VALUE CONCEPTS

By making use of the undefined generic value concept expressed by '*p* is better than *q*' and the concepts we have just defined in terms of it, we may define the concept of *intrinsic preferability* – the concept of one state of affairs being *intrinsically better* than another. Then by reference to intrinsic preferability, we may define *intrinsic goodness*, *intrinsic badness* and *intrinsic neutrality.*

The concept of intrinsic preferability, then, may be defined as follows:

---

affairs, *p* and *q*, *p* entails the disjunction, *p* or *q*. Compare the undefined concept of 'evaluatively basic proposition,' introduced in Warren S. Quinn, 'Theories of Intrinsic Value,' *American Philosophical Quarterly,* XI (1974), 123-132; see pp. 128ff.

> p is intrinsically better than q = *Df* p and q are bearers of intrinsic value; and any world *W*1, such that p reflects all the good and evil there is in *W*1, is better than any world *W*2 such that q reflects all the good and evil there is in *W*2.

The second clause of the definiendum tells us the sense in which the result of considering p and q 'in isolation' enables us to say that p is intrinsically better than q.

The only states of affairs that are terms of the intrinsic preferability relation will be intrinsic value states.[6] Thus the mixed disjunction, *there are three happy Canadians or there are three happy non-Canadians* (p), is not intrinsically better than *there are stones* (q); for p is not an intrinsic value state (it cannot contain all the good and evil that there is in any possible world), and hence it is not itself a bearer of intrinsic value.

Given this concept of intrinsic preferability, we may now explicate the various intrinsic value concepts and exhibit the relations among them. To illustrate one way of doing this, I will adapt certain features of the system of intrinsic value that has been set forth by Ernest Sosa and me.[7]

Abbreviating 'p is intrinsically preferable to q' as 'pPq', let us say:

> p has the same intrinsic value as q (pSq) = *Df* p and q are bearers of intrinsic value; and $\neg(pPq)$ and $\neg(qPp)$.
>
> p is intrinsically indifferent (Ip) = *Df* $pS\neg p$.
>
> p is intrinsically neutral (Np) = *Df* $(\exists q)(Iq$ and $pSq)$.
>
> p is intrinsically good (Gp) = *Df* $(\exists q)(Iq$ and $pPq)$.
>
> p is intrisically bad (Bp) = *Df* $(\exists q)(Iq$ and $qPp)$.

These definitions may be put informally as follows (provided the qualification 'intrinsically' is understood throughout): one state of affairs has the *same intrinsic value* as another if and only if both are bearers of intrinsic value and neither one is intrinsically better than the other; the *indifferent* is that which has the same value as

[6] An alternative to the present procedure would be to characterize the intrinsic value of a state of affairs p as being a function of the value of those intrinsic value states that reflect all the good and evil that there is in p. Then all states of affairs, and not just the ones I have called 'bearers of value,' would fall within the field of the intrinsic preferability relation. But the present procedure, as we shall see, enables us to solve the problem of the intrinsic value in disjunctive states of affairs. And the alternative does not have this advantage.

[7] Roderick M. Chisholm and Ernest Sosa, 'On the Logic of "Intrinsically Better",' *American Philosophical Quarterly,* III (1966), 244-9. In this system we made use of rules corresponding to *modus ponens* and to a principle of substitution for logically equivalent states of affairs. The latter principle enabled us to say that all logically equivalent states of affairs are the same in value. But that system, unlike the one presented here, did not restrict the intrinsic preferability relation to those states of affairs that are here called the bearers of intrinsic value.

its negation; the *neutral* is that which has the same intrinsic value as the indifferent; the *good* is that which is intrinsically better than the indifferent; and the *bad* is that which is such that the indifferent is intrinsically better than it.

I add this further definition to facilitate exposition:

> p is at least as good intrinsically as q (pAq) = *Df* p and q are bearers of intrinsic value; and $\neg(qPp)$.

In other words, if p and q are bearers of intrinsic value, and if q is not intrinsically better than p, then p is *at least as good* intrinsically as q.

The axioms of the system may now be abbreviated as follows:

A1 $(p)\,(q)\,[pPq \rightarrow \neg(qPp)]$.

A2 $(p)\,(q)\,(r)\,[(qAp \text{ and } rAq) \rightarrow rAp]$.

A3 $(p)\,(q)\,[(Ip \text{ and } Iq) \rightarrow pSq]$.

A4 $(p)\,(Gp \vee B\neg p) \rightarrow pP\neg p$.

In other words: (1) if p is better than q, then q is not better than p; (2) if q is at least as good as p, and if r is at least as good as q, then r is at least as good as p; (3) if p is indifferent and q is indifferent, then p has the same value as q; and (4) if p is good or if not-p is bad, then p is better than not-p.

It is important to note the distinction between intrinsic indifference (Ip) and intrinsic neutrality (Np). The class of the neutral is wider than that of the indifferent. A neutral state of affairs (e.g., *there being no pleasure*) may have a good negation. And a neutral state of affairs (e.g., *there being no displeasure*) may have a bad negation. But an indifferent state of affairs (e.g., *there being stones*) is a neutral state of affairs that has a neutral negation.

According to some theories of intrinsic value, but not according to the present theory, the negation of a state of affairs that is intrinsically good is intrinsically bad, and the negation of a state of affairs that is intrinsically bad is intrinsically good. Our intuition, however, may be more nearly in accord with that of Aristotle: "Positive goodness and badness are more important than the mere absence of goodness and badness: for positive goodness and badness are ends, which the mere absence of them cannot be."[8] Thus, if we are hedonists, we will say that *someone experiencing pleasure* is intrinsically good. But we will not say that its negation – *no one experiencing pleasure* – is intrinsically bad. For the negation may obtain in worlds in which nothing is good and nothing is bad. Thus it may obtain in worlds in which there are no living things and in which, therefore, there is neither pleasure nor displeasure. Analogous points may be made with respect to other things that have been said to be intrinsically bad or intrinsically good.

---

[8] *Rhetoric,* Book I, Ch. 7, 1364a.

The present conception of the relation between good and evil was put in the following way by Oskar Kraus: “The nonexistence of a good is not an evil, and the nonexistence of an evil is not a good; one can say only that the existence of a good is preferable to its nonexistence, and conversely in the case of evil.”[9]

5. THE INTRINSIC VALUE IN COMPOUND STATES OF AFFAIRS

Sosa and I did not set forth any axioms pertaining to the intrinsic value to be found in disjunctive states of affairs. The present approach to intrinsic value may now be seen to justify this omission.[10] Let us consider, then, the general question of the intrinsic value to be found in compound states of affairs.

In order to make certain theoretical points, I will suppose once again that hedonism is the correct theory of value. In other words, I will suppose that the only things that are intrinsically good are states of affairs implying that there is pleasure and that the only things that are intrinsically bad are states of affairs implying that there is displeasure. I will also suppose that pleasures and displeasures can be ordinally ranked, that the more pleasure the better and the less displeasure the better, and that equal amounts of pleasure and displeasure balance each other off. (But these hedonistic assumptions are not at all essential to the general conception of value that is here defended. They are made only for the purpose of simplifying examples.)

Let us now consider the following states of affairs, all being, according to our present account, bearers of value:

(p) Jones experiencing 1 unit of pleasure.

(q) Smith experiencing 1 unit of displeasure.

(r) There being stones.

(s) Brown experiencing 2 units of pleasure.

(t) Black experiencing 2 units of displeasure.

Given our hedonistic assumptions, there is, of course, no problem in assessing the intrinsic value of these five states of affairs and in ranking them. Thus p and s are good, q and t are bad, r is neutral, s is the best of the lot, and t is the worst.

---

[9] Oskar Kraus, *Die Werttheorien: Geschichte und Kritik* (Brunn: Verlag Rudolf M. Rohrer 1937) 227. Compare Georg Katkov, *Untersuchungen zur Werttheorie und Theodizee* (Brunn: Verlag Rudolf M. Rohrer 1937), 67ff.

[10] Concerning the intrinsic value of disjunctions, compare Lennart Åqvist, ‘Chisholm-Sosa Logics of Intrinsic Betterness and Value,’ *Nous*, II (1968), 253-70; and Roderick M. Chisholm, ‘The Intrinsic Value in Disjunctive States of Affairs,’ *Nous*, IX (1975), 295-308. The present approach would render superfluous the axioms about disjunction that were defended in the last-named paper.

There is no theoretical problem involved in evaluating the various possible *conjunctions* of the above states of affairs. Thus *p* and *q* is neutral; *p* and *r* is good; *p* and *s* is good; and *p* and *t* is bad and has the same value as *q*... And the conjunction of all five is neutral.

What of the *negations* of these states of affairs?

According to many theories of value, as we have noted, the negation of a bad state of affairs is good, and the negation of a good state of affairs is bad. According to our present criteria, however, the negation of each of these states of affairs is neutral. For each negation may obtain in worlds in which there is no pleasure and no displeasure, and hence in worlds in which (according to our hedonistic assumptions) there is nothing that is intrinsically good and nothing that is intrinsically bad. Any world which is such that all the good and evil that there is in that world may be found in the negations of any of the above states of affairs will be a world that is intrinsically neutral.

And now what of the *disjunctions* of these states of affairs? No disjunction, having any two of the above states of affairs as its disjuncts, is a bearer of intrinsic value. For none is capable of reflecting all the good and evil that there is in any possible world. Consider, for example, $p \vee q$. Any world in which $p \vee q$ obtains will have, *in addition* to whatever generic value there might be in $p \vee q$, *also* either the positive intrinsic value that is in *p* or the negative intrinsic value that is in *q* or both. And analogously for the other disjunctions that may be formed from the above. None of them is a bearer of intrinsic value and hence none of them need be considered in assessing the value of any world in which it obtains.

Castañeda has noted that three quite different procedures have been used in evaluating disjunctions. These may be characterized by the following three rules: (a) the value of the disjunction is the same as that of the higher valued disjunct; (b) the value of the disjunction is the same as the value of the disjuncts if these are the same in value, and otherwise it lies between the disjuncts in value; and (c) the value of the disjunction is the same as the value of its lower valued disjunct.[11] These three procedures would have very different results in application to the disjunctions we have been considering. Consider only $p \vee q$. According to (a) it would be good; according to (b) it would be neutral; and according to (c) it would be bad. Hence these procedures would seem to be fundamentally different. And so we may feel that we should select just one of them. But how are we to make the choice?

The reasonable thing is to assume that they are *all* wrong.

Different as the three procedures may be in application to restricted states of affairs, they do not differ at all in application to possible worlds. That is to say, when we come to evaluate worlds, it will not matter in the least which method we choose, for the results will be the same in each of the three cases. If we know the value of the disjuncts of any given disjunction, and if we know which of the disjuncts obtains, then there is no point in trying to calculate the value of the

[11] See Hector-Neri Castañeda, 'Ought, Value, and Utilitarianism,' *American Philosophical Quarterly*, VI (1969), 257-275.

disjunction. Any dispute, then, about the intrinsic value of the disjunctions of our five states of affairs above would seem to be entirely idle. Hence the finding that they have no intrinsic value at all should be one that is entirely welcome. And this finding is one of the consequences of the present explication of intrinsic value.

CHAPTER 2

E. BODANSZKY AND E. CONEE

ISOLATING INTRINSIC VALUE

Roderick Chisholm has recently proposed analyses of the basic concepts of intrinsic value. ('Intrinsic Value', *Values and Morals*, ed. A. I. Goldman and J. Kim, Reidel 1978, pp. 121-30 [* pp. 1-10 of this volume]). He derives his approach from the view of Aristotle and Moore that the intrinsic value of something is the value that it "would have in isolation". Chisholm seeks to defend this view by offering precise analyses which incorporate the intuition that an intrinsically good state of affairs is one such that 'its goodness doesn't require that there obtain some *other* good state of affairs which neither includes nor is included within it' (ibid. pp. 121-2 [* p. 2 of this volume]). We shall try to show here that Chisholm's proposed analyses do not succeed. We maintain that they fail in ways that indicate that nothing like his isolation approach yields adequate explanations of these concepts.

Chisholm's account appeals to a single normative primitive: the concept of being "generically better". No explanation of this notion is given. The way it is used in the exposition of the account requires that it establish a value ranking in which a state of affairs gets credit for supplying either intrinsic or extrinsic benefits, and suffers discredit for having either intrinsic or extrinsic liabilities. For example, something intrinsically neutral is generically better than neutral if it has good effects, and something intrinsically good is generically neutral if the disvalue of its effects is equal to its intrinsic value (see ibid. pp. 121, 122, 124 [* pp. 2-4 of this volume]). Appealing to this notion of being generically better, Chisholm constructs his analysis of being intrinsically better, and uses that to explain intrinsic goodness and evil. This is the analysis:

D1 *p is intrinsically better than q* $=_{df}$ p and q are bearers of intrinsic value; and any world, W_1, such that p reflects all the good and evil there is in W_1, is [generically] better than any world, W_2, such that q reflects all the good and evil there is in W_2 (ibid p. 126 [* p. 6 of this volume]).

The technical terms of D_1 are defined as follows:

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 11-13.
© 2005 *Springer. Printed in the Netherlands.*

D2 *p reflects all the good and evil there is in q* $=_{df}$ q entails p; and every part of q that entails p has the same [generic] value as p (ibid p. 123 [* p. 3 of this volume]).

The relevant concepts of entailment and parthood are these: 'p entails q $=_{df}$ p is necessarily such that (i) if it obtains then q obtains and (ii) whoever accepts it accepts q, . . . a state of affairs p is a "part" of a state of affairs q, if and only if q entails p' (ibid. [* p. 3 of this volume]).

D3 *p is an intrinsic value state* $=_{df}$ there is a possible world, W, such that: p reflects all the good and evil there is in W; and if p is not [generically] neutral, then everything that reflects all the good and evil there is in W entails p or is entailed by p (ibid. p. 124 [* p. 4 of this volume]).

D4 *p is a bearer of intrinsic value* $=_{df}$ p is an intrinsic value state and not-p is an intrinsic value state (ibid. p. 125 [* p. 5 of this volume]).

Unfortunately, even the elaborate conceptual machinery of D1 does not suffice to "isolate" the intrinsic goods. They retain logical and causal accompaniments that ruin the account. In the first place, most of the examples of intrinsic goods that Chisholm uses in developing the analyses fail to be bearers of intrinsic value, although by D1 only bearers of intrinsic value are intrinsically better than anything. Consider:

(a) Brown experiencing two units of pleasure.

For (a) to be a bearer of intrinsic value, it must be an intrinsic value state. It is not. The worlds where (a) reflects all the good and evil must be worlds in which the only thing that independently contributes value is Brown's pleasure, if D1 is to work properly. But in any such world there are many logically independent states of affairs which also reflect all the good and evil, for instance either

(b) Someone experiencing two units of pleasure while wearing a hat,

or

(b') Someone experiencing two units of pleasure while not wearing a hat.

So each such world contains a part (either (b) or (b')) which reflects all the good and evil there, but neither entails nor is entailed by (a). Since (a) is non-neutral in these worlds, by D3 it is not an intrinsic value state. Thus D4 tells us that it is not a bearer of intrinsic value, and so not counted by D1 as intrinsically better than anything.

The causal problem for D1 is made by the fact that intrinsically good states of affairs that can have bad effects reflect all the good and evil in some generically neutral worlds. Because of the way D1 evaluates states by use of such worlds, these fail to be rated intrinsically better than neutral (Chisholm was aware that this problem might arise, and sought to avoid it – see ibid. p. 124 [* p. 4 of this volume]; we argue that the problem remains). In the case of (a), this can be shown as follows. There is a possible world, Wa, where (a) causes

(c) Brown experiencing two units of *dis*pleasure,

and those two experiences are the only two things that independently contribute value to the world. Nonetheless, (a) reflects all the good and evil in Wa. Since (a) causes the intrinsically evil (c), it has neutral generic value in Wa. The other parts of Wa that entail (a) (including Wa itself) are conjunctions of (a) with the other truths of Wa. When such a part also includes (c) as a conjunct, it is clearly intrinsically and generically neutral. But even when such a part does not include (c) it is generically neutral, for it is (a)'s *generic* value in Wa that determines its contribution to the generic value of these conjunctions. To see this, consider what would be relevant to the evaluation of such a conjunction if it were the only known consequence of a contemplated act. We would want to know its generic value, since we would be concerned with both its intrinsic and extrinsic contributions. And it would be a clear mistake to ignore (a)'s extrinsic disvalue in making this assessment. That would rate too highly the act of bringing about the conjunction. Thus the generic value of a conjunction depends upon that of its conjuncts. So the parts of Wa that entail (a) are all generically neutral. Since Wa is a world where (a) reflects all the good and evil, according to D1 (a) is not intrinsically better than intrinsically neutral states of affairs.

We believe that D1 is beyond repair. And we think that its problems typify the difficulties that face attempts to characterise intrinsic betterness by use of generic betterness and some sort of isolation. In particular, reflection upon these problems makes it seem quite doubtful that there is any sort of situation where each state of affairs has exactly its intrinsic value as its generic value. And without that there appears to be no hope for the approach.

CHAPTER 3

R. M. CHISHOLM

DEFINING INTRINSIC VALUE

In their paper, 'Isolating Intrinsic Value', (*Analysis*, January 1981, pp. 51-3 [* pp. 11-13 of this volume]) Bodanszky and Conee express doubts as to whether an "isolation approach" will yield adequate definitions of intrinsic value and related concepts. I think that they are right in their criticism of the particular attempt that they discuss. But I am quite certain that they are wrong in concluding that no such "isolation approach" yields adequate explications of these concepts.

I believe they assume, as I do, that the bearers of intrinsic value are states of affairs – those abstract objects, designated by that-clauses and sentential gerundives, which are sometimes called "propositions". *Someone experiencing pleasure*, for example, would be a state of affairs that is intrinsically good and *someone feeling displeasure* a state of affairs that is intrinsically bad.

Perhaps they also agree with me that the intrinsic value of a state of affairs, as contrasted with its instrumental value, is a feature that that state of affairs has necessarily. If a state of affairs is intrinsically good, then it is intrinsically good in every possible world in which it obtains (or is true). But a state of affairs that is instrumentally good need not be instrumentally good in every possible world in which it obtains (or is true). And analogously for badness and indifference.

It can be shown, I think, that all intrinsic value concepts may be analysed in terms of intrinsic preferability. In the definitions that follow 'intrinsically' should be thought of as implicit throughout.

A state of affairs p is *the same in value* as a state of affairs q, if and only if p is not preferable to q and q is not preferable to p. A state of affairs is *indifferent* if and only if it is the same in value as its negation. A state of affairs is *good* if and only if it is preferable to a state of affairs that is indifferent. And a state of affairs is *bad* if and only if a state of affairs that is indifferent is preferable to it (compare Roderick M. Chisholm and Ernest Sosa, 'On the Logic of "Intrinsically Better",' *American Philosophical Quarterly*, Vol. III, 1966, pp. 244-49).

The question, then, turns upon the analysis of the locution 'p is intrinsically preferable to q'. What can we say about it?

Following Brentano and Meinong, I would suggest that intrinsic value concepts may be defined in terms of the appropriateness of certain intentional attitudes. For certain attitudes may be said to be *appropriate to* – and indeed *required by* – their

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 15-16.
© 2005 *Springer. Printed in the Netherlands.*

objects (compare Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, London: Routledge & Kegan Paul, 1969, and Alexius Meinong 'Zur Grundlegung der allgemeinen Werttheorie', and 'Ethische Bausteine', in *Abhandlungen zur Werttheorie*, in *Meinong Gesamt Ausgabe*, Vol. III, Graz, Akademische Druck- u.Verlagsanstalt, 1968). Thus a pro-attitude is appropriate to a state affairs that is intrinsically good, and an anti-attitude is appropriate to a state of affairs that is intrinsically bad. How, then, are we to define intrinsic preferability?

If a state of affairs p is intrinsically better than a state of affairs q, then it is appropriate to *prefer* p to q. And the contemplation of the two *requires* that p be preferred to q. Hence we could explicate intrinsic preferability this way:

> p is intrinsically preferable to q $=_{df}$ p and q are necessarily such that, for any x, the contemplation of just p and q by x requires that x prefer p to q.

One might qualify the definition by saying: '…the contemplation of just p and q *as such* by x requires that x prefer p *as such* to q.' One would thus be using the 'as such' to stress the fact that the contemplation of a state of affairs *as such* is the contemplation just of *that* state of affairs – as distinguished, for example, from the contemplation of some wider state of affairs which one may think that the given state of affairs brings along with it. And one may say 'x prefers p as such to q' in order to stress the fact that x is ranking just *those* states of affairs, as distinguished from any wider states of affairs that may include them. But I will assume that the reader takes the definition literally: 'the contemplation of just p and q' would not refer to the contemplation of any wider state of affairs that might include p or q; and 'x prefers p to q' would not refer to a ranking of any states of affairs that are wider than p or q.

It may be noted that, in ranking p and q as such and thus 'in isolation,' one is concerned just with those features of p and q that they have necessarily and not with any particular features that they may happen to have in this world but not in others. And so one might say that the states of affairs are objects of the "antecedent will" as distinguished from the "consequent will".

This approach to intrinsic value, then, makes use of the concept of *requirement.* And the concept of requirement, there is reason to think, is the central concept of ethics (compare Roderick M. Chisholm, 'Practical Reason and the Logic of Requirement', in Stephan Körner, ed., *Practical Reason*, Oxford, Basil Blackwell, 1974; and Philip C. Quinn, *Divine Commands and Moral Requirements*, Oxford, The Clarendon Press, 1978). The "isolation approach", therefore, not only yields adequate definitions of the basic intrinsic value concepts, but has the following advantage as well: it provides a way of reducing the concepts of the theory of value ("axiology") to those of ethics ("deontology").

CHAPTER 4

N. M. LEMOS

THE CONCEPT OF INTRINSIC VALUE

My main concern in this chapter is to explicate the concept of intrinsic value. I discuss and defend the view that the concept of intrinsic value may be explicated in terms of the concept of "correct" or "required" emotion. I am not especially concerned with whether this explication amounts to a definition or philosophical analysis of the concept of intrinsic value, nor am I especially interested in "reducing" the concept of intrinsic value to certain other concepts. I am simply concerned with explaining what I take intrinsic value to be or, alternatively, what it is for something to be intrinsically valuable.

I wish to begin, however, by describing certain general views belonging to one traditional way of thinking about intrinsic value. These views are among the main theses of a tradition whose representatives include Franz Brentano, G. E. Moore, W. D. Ross, and A. C. Ewing. In stating these general views, I shall be describing, in part, the core of this tradition. I do this for two reasons. First, though I shall not undertake to defend them in this chapter, I think these theses pertaining to the nature and concept of intrinsic value are both plausible and true. Second, and more important, these remarks will provide some general background against which the explication of intrinsic value may proceed. It is hoped that these remarks will help illustrate in rough outline the concept with which I am concerned.

First, the traditional view holds that if something is intrinsically good, it is not intrinsically bad or intrinsically neutral or indifferent; and if something is intrinsically bad, it is not intrinsically good or indifferent. According to the traditional view, the claims that (1) *X* is intrinsically good and (2) *X* is intrinsically bad are contraries. Similarly, the traditional view assumes that if *X* is intrinsically better than *Y,* then it is false that *Y* is intrinsically better than *X* or that they are the same in intrinsic value.

The second feature is difficult to state precisely, but let us say that according to the traditional view, intrinsic value is a *nonrelational* concept. When one says that something is intrinsically good, in the sense with which we are concerned, he means just that, that it is intrinsically good *period.* He does not mean that it is intrinsically good for me, for himself, for human beings, or for rational beings. In this respect,

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 17-31.
© 2005 *Springer. Printed in the Netherlands.*

claims that something has intrinsic value are not like the claim "Boston is near," for in ordinary contexts the latter claim is meaningful only if one takes the speaker to be saying that Boston is near to something (e.g., to the speaker, to the person addressed, to Providence). If it is not clear what the speaker is saying Boston is near to, it is appropriate to ask, "Boston is near to what?" The tradition with which I am concerned does not take the concept of intrinsic value to be a relational concept, relative to persons, species, or anything else. If someone insists that he is concerned with such a relational concept or with what it is for something to be intrinsically good for someone, then he is talking about something other than the traditional concept of intrinsic value with which I am concerned.

Even if intrinsic value is not a relational concept, one can still talk meaningfully of something's being good for one person and not for another. One may say, for example, that insulin is good for a diabetic but not for an otherwise healthy person. But clearly, what this means is not that insulin is intrinsically good for one person and not for another. What is meant is roughly that insulin has an instrumental value when taken by someone with diabetes but not when taken by an otherwise healthy person. Furthermore, we may also say, for example, that a state of affairs is intrinsically better for one person than for another. We might say that the state of affairs consisting in Smith's being very happy and Brown's being very unhappy is intrinsically better for Smith than for Brown. But this may be taken to mean roughly that Smith and Brown have certain characteristics, *F* and *G*, such that Smith's having *F* is intrinsically better than Brown's having *G*. The fact that one person is better off than another when certain states of affairs obtain does not imply that intrinsic value is a relational concept.

Third, the philosophers in this tradition hold that we *know* that some things are intrinsically good, that some things are intrinsically bad, and that some things are intrinsically better than others. They are *cognitivists* about intrinsic value. There are significant differences among them concerning the nature of our knowledge of value and differences about how we have such knowledge, but they all agree that we know that some things are intrinsically valuable. They do not maintain, however, that we know or can know, upon reflection or investigation, the answer to every question of the form "Is *X* intrinsically good?" or "Is *X* intrinsically better than *Y*?" Cognitivism about value does not imply or require omniscience about value. Withholding belief may be the epistemically reasonable stance to take with respect to certain claims about intrinsic value. Moreover, since knowledge implies true belief, this stance implies that it is true that some things are intrinsically good, others are intrinsically bad, and some are intrinsically better than others. There are truths about intrinsic value such that whoever denies them denies what is true.

The fourth feature of this tradition is also hard to state precisely, but we may say roughly that it takes intrinsic value to be *distinct* from any "natural" property, relation, or state of affairs. The philosophers in this tradition assume that we can distinguish ethical properties, relations, or states of affairs from natural entities. Intrinsic value is distinct from any natural entity in the sense that it is not identical with any such thing. Thus, if there is a property of being intrinsically good, that

property is not identical with any natural property, such as being pleasant or being desired by most people. Similarly, if there is a state of affairs that is expressed by the sentence "John's being happy is intrinsically good," then the state of affairs expressed by that sentence is not identical with that expressed by the sentences "John's being happy would be approved by most people" or "John desires upon reflection that he be happy."

Although intrinsic value is not identical with any natural entity, many of the members of this tradition also hold that intrinsic value is dependent on nonethical features or facts, that whatever has intrinsic value has it in virtue of or because of the nonethical properties of that thing or in virtue of the obtaining of nonethical states of affairs. In this respect, they assume that intrinsic value is like certain other evaluative or normative concepts. Just as the beauty of a painting may be thought to depend on its colors and their arrangement or the goodness of an apple on its being sweet, juicy, ripe, and so on, intrinsic value may be thought to depend on certain nonethical facts or properties. The claim that intrinsic value is distinct from the natural must not, therefore, be taken to imply that intrinsic value does not supervene on or is not determined by the nonethical. It should not be taken to deny that intrinsic value is intimately related to the nonethical.

Finally, according to this tradition, the intrinsic value of a thing is not dependent on its being the object of any psychological attitude. If a thing has intrinsic value, it has it independently of its being the object of any psychological attitude or its being conducive to or productive of any such attitude. If a thing has intrinsic value, it does not have that value because, or in virtue of, its being the object of anyone's psychological attitude or because it would be the object of such an attitude under some set of hypothetical conditions. We may also express this claim by saying that the intrinsic value that a thing possesses is not conferred on it by its being the object of such attitudes. Thus, if *X* is intrinsically good, then *X* is not intrinsically good because or in virtue of the fact that someone or some group likes or desires *X* or would like *X* if they were, say, "fully informed." This is an essentially negative view. It simply tells us what does not confer intrinsic value, what does not make things intrinsically valuable. It is important not to misunderstand this claim. Some writers in this tradition have held that if anything is intrinsically good, then that thing must have or contain consciousness as a part. Moore, for example, changed the position taken in *Principia Ethica* and writes in the later work, *Ethics*, that "it does seem as if nothing can be an intrinsic good, unless it contains both some feeling and also some other form of consciousness."[1] We must distinguish, then, between something's having consciousness as a part or implying that something is conscious from something's being the object of consciousness. It is possible that something have the first characteristic without having the second. For example, the state of affairs "there being at least 1,248 conscious beings" necessarily implies that something is conscious, but that state of affairs need not have been itself the object of anyone's psychological attitudes or consciousness. If we make this distinction,

---

[1] G. E. Moore, *Ethics* (Oxford: Oxford University Press, 1912), p 107.

then we might consistently assert that every intrinsically good whole has consciousness as a part while denying that anything is intrinsically good because or in virtue of its being the object of consciousness.

1. CORRECT EMOTION AND INTRINSIC VALUE

What is it for something to be intrinsically valuable? What is it for something to be intrinsically good or bad or better than something else? One traditional way of answering these questions seeks to explicate the concept of intrinsic value in terms of the concept of correct, suitable, or fitting emotion. Among those favoring this approach are Franz Brentano, C. D. Broad, A. C. Ewing, and Roderick Chisholm. In spite of certain significant differences, these writers hold that something's being intrinsically good may be understood in terms of its being correct or fitting to love or like that thing in and for itself or for its own sake. Similarly, *A*'s being intrinsically better than *B* may be understood in terms of the correctness or fittingness of preferring *A* in itself to *B*. Brentano, for example, says, "We call a thing good when the love relating to it is correct. In the broadest sense of the term, the good is that which is worthy of love, that which can be loved with a love that is correct."[2] He adds, "One loves or hates correctly provided that one's feelings are adequate to their object – adequate in the sense of being appropriate, suitable, or fitting."[3] And again, "when we call certain objects good and others bad we are merely saying that whoever loves the former and hates the latter has taken the right stand."[4]

This approach to explicating the concept of intrinsic value has a certain intuitive appeal. It is plausible to think of the good as being worthy of, or meriting, love or favor and the bad as being worthy of, or meriting, dislike or disfavor. There thus seems to be at least this much truth in Brentano's remark – that whoever loves what is good and hates what is bad has taken the right stand. But what is it to take the right stand? We may say that it is a matter of loving, hating, or preferring fittingly. Whoever loves in itself what is intrinsically good loves that thing fittingly. It is plausible to think that there is some way in which the concept of value is related to our emotional attitudes, to "pro attitudes" and "anti attitudes." Although this approach does not tell us that such attitudes confer value on things or that value consists in our having such attitudes toward things, it does tell us that there is some sort of normative connection between the things that have value and our favoring and disfavoring.

The view that our emotional attitudes can be ethically appropriate or fitting to things has an ancient and venerable history reaching back at least as far as

---

[2] Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, English edition edited by Roderick Chisholm and translated by Roderick Chisholm and Elizabeth Schneewind (London: Routledge & Kegan Paul, 1969), p. 18.
[3] Ibid., p. 74.
[4] Franz Brentano, The *Foundation and Construction of Ethics*, English edition edited and translated by Elizabeth Hughes Schneewind (New York: Humanities Press, 1973), p. 131.

Aristotle's claim that there is a mean for emotion and feeling, as well as for action. For example, we find Aristotle claiming that there is a mean in anger, that anger toward certain forms of injustice is appropriate, whereas envy and spite consist in having inappropriate feelings of hatred or pleasure. He writes, "the man who is characterized by righteous indignation is pained at undeserved good fortune, the envious man, going beyond him, is pained at all good fortune, and the spiteful man falls so far short of being pained that he even rejoices [when someone suffers]."[5] According to Aristotle, it is ethically unfitting or inappropriate to take pleasure in the suffering of others and unfitting to be pained or displeased at the good fortune of those who merit good fortune. A similar view is expressed by Schopenhauer: "In a certain sense the opposite of envy is the habit of gloating over the misfortunes of others. At any rate, whereas the former is human, the latter is diabolical. There is no sign more infallible of an entirely bad heart, and of profound moral worthlessness than open and candid enjoyment of seeing other people suffer."[6] In addition to these examples, we may add that fear is fitting to certain sorts of dangers; gratitude toward certain sorts of favors; admiration toward excellences such as wisdom, courage, and aesthetic virtuosity; and remorse and guilt toward one's own wrongdoing. The list could, of course, go on, and one would be hard pressed to find a better study of the variety of fitting or appropriate attitudes and emotion than Aristotle's *Ethics.*

If the concept of intrinsic value can be explicated in terms of correct or fitting emotion, what form will the explication take? Some attempts seem pretty clearly unsatisfactory. Suppose, for example, we say, "*X* is intrinsically good just in case *X* is fittingly or correctly loved for its own sake." This view is unacceptable because it implies that the only things that are intrinsically good are those that are in fact loved. It is not at all clear that this is true, and it should not be a logical consequence of our explication of intrinsic value. Consider also the claim that "*X* is intrinsically good just in case if *X* is loved for its own sake, then *X* is correctly or fittingly loved." If the explanans is understood in terms of material implication, then one unhappy consequence is that anything that is not loved for its own sake is intrinsically good, and that is surely false.

Rather than consider all attempts to explicate the concept of intrinsic value in terms of correct emotion, a task that would be tediously long, I shall focus on one promising line of approach. C. D. Broad once suggested that "*X* is good" can "be defined as meaning that *X* is such that it would be a fitting object of desire to any mind which had an adequate idea of its non-ethical characteristics."[7] More recently, Roderick Chisholm has claimed that "*p* is intrinsically better than *q*" can be defined

---

[5] Aristotle, *Nicomachean Ethics*, translated by W. D. Ross in *The Basic Works of Aristotle*, edited by Richard McKeon (New York: Random House, 1941) Bk. II, Chapter 7, 1108b. Compare also Aristotle's remarks on pity and indignation in *Rhetoric*, translated by W. Rhys Roberts Bk. 11, Chapter 9, 1386b.

[6] Arthur Schopenhauer, *The Basis of Morality*, translated by A. Broderick Bullock (London: Swan Sonneschein, 1903), pp. 156-7.

[7] C. D. Broad, *Five Types of Ethical Theory* (New York: Harcourt, Brace and Co., 1930), p. 283.

as follows: "*p* and *q* are necessarily such that, for any *x*, the contemplation of just *p* and *q* by *x* requires that *x* prefer *p* to *q*."[8]

Broad's definition involves one's having "an adequate idea" of something, and Chisholm's refers to the contemplation of just *p* and *q*. Why not say more simply "*p* and *q* are necessarily such that, for any *x*, *p* and *q* require that *x* prefer *p* to *q*"? One reason to favor the more complex formulation is that *p* and *q* might be very complicated states of affairs, too complex for some persons to conceive or contemplate, or *p* and *q* might involve concepts or properties that some persons cannot grasp. If there are states of affairs that some persons are incapable of conceiving, it is not clear that they can be required to have a pro-attitude toward them or that they can be required to prefer one such state of affairs to another.

In attempting to explicate the concept of intrinsic value in terms of correct or fitting emotion, one must say something about the sorts of emotional attitudes involved. Chisholm says:

> One might qualify the definition by saying "... the contemplation of just *p* and *q* as such by *x* requires that *x* prefer *p* as such to *q*." One would thus be using the "as such" to stress the fact that the contemplation of a state of affairs *as such* is the contemplation just of *that* state of affairs − as distinguished, for example, from the contemplation of some wider state of affairs which one may think that the given state of affairs brings along with it.[9]

Analogously, one could say that *x* is required to prefer *p* as such to *q* to emphasize that *x* is required to rank just those states of affairs, as distinguished from any wider states of affairs that might include them or any other states of affairs that might be among their causal consequences.

Chisholm's emphasis on preferring *p* as such to *q* reflects a distinction drawn by Brentano concerning our emotional attitudes. Within the sphere of emotional phenomena, Brentano distinguishes three basic types: love simpliciter, hate simpliciter, and preference simpliciter. To love something simpliciter is to love that thing *as such*, to feel favorably toward a thing *in and for itself.* Similarly, to hate something simpliciter is to hate that thing as such, to have an anti feeling toward it in and for itself. To prefer simpliciter *p* to *q* is to prefer *p* as such or in and for itself to *q*.

Love, hate, and preference simpliciter must be distinguished from loving, hating, or preferring something (1) as a means, (2) from choosing, and (3) having such attitudes *per accidens.* One may prefer taking aspirin to taking a sugar pill as a means of relieving a headache, but it does not follow that one prefers taking aspirin as such or in itself to taking a sugar pill. Preference simpliciter must also be distinguished from choice. If one chooses to bring about *A* rather than *B,* it does not follow that one prefers *A* in itself to *B.* Choosing to take an aspirin rather than a sugar pill does not imply that one prefers the former as such to the latter any more

---

[8] Roderick Chisholm, "Defining Intrinsic Value," *Analysis* 41, (March 1981), p. 100 [* p. 16 of this volume].
[9] Ibid., p. 100 [* p. 16 of this volume].

than choosing to spend a night at the movies rather than at the opera implies that one prefers the former in itself to the latter. Whether it is appropriate to choose *A* over *B* typically involves consideration of states of affairs that are wider than or different from *A* and *B* themselves, such as those states of affairs that are the consequences of bringing about *A* and those of bringing about *B.* Whether it is appropriate to choose a night at the movies over a night at the opera depends in part on the cost and consequences of doing the one or the other. Finally, loving, hating, and preferring simpliciter must be distinguished from having such attitudes *per accidens.* If one loves simpliciter a certain whole, then one may be said to love the parts of that whole *per accidens.* However, it does not follow that one loves every part of that whole simpliciter. Thus, one can love simpliciter Smith's overcoming temptation and one can love simpliciter Brown's remorse toward his wrongdoing, but it does not follow that one loves simpliciter Smith's being tempted or Brown's doing wrong. Similarly, one might hate simpliciter Smith's taking joy in his misdeeds, but it does not follow that one hates Smith's having joy as such.

It might be useful to compare the sort of approach we have been discussing with that adopted by Moore and Ross. Insofar as Chisholm stresses the notion that the contemplation of just *p* and *q* requires that one prefer *p* as such to *q,* there is some justification for saying that he adopts an "isolation approach" in explicating the concept of intrinsic value. We may say that his approach is *intentionally isolationist* because it stresses the intentional attitudes of contemplating and preferring states of affairs as such, in isolation from the contemplation and ranking of other, wider states of affairs. We may contrast this form of isolation approach with what we may call "ontological isolationism." Some writers, such as Moore and Ross, have suggested that "by calling a thing intrinsically good we mean that it would be good even if nothing else existed."[10] This sort of ontological isolationism is not very helpful since there are certain sorts of things that are intrinsically good but simply could not be the only things that exist. Consider the fact of Smith's being happy and let's suppose that it is intrinsically good. If there are certain abstract entities such as numbers or properties or states of affairs that necessarily exist, it would be impossible for Smith's being happy to be the only thing that exists. More important, though, is the fact that Smith's being happy could not exist without Smith's existing, as well as, I suppose, Smith's having certain pleasures and certain desires satisfied and his having certain beliefs to the effect that he had those pleasures and that his desires were satisfied. Since it is necessarily false that Smith's being happy could be the only existing thing, this sort of ontological isolationism is not very clear or very helpful.

We should note that, in spite of these different approaches, both Chisholm and Moore hold that if a thing has a certain intrinsic value, then it must have that value whenever it occurs. This thesis, sometimes referred to as the "thesis of universality," is implied by Chisholm's definition of intrinsic value. If *p*'s being

[10] Moore, *Ethics*, p. 38; W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930), p. 73.

intrinsically better than *q* is a matter of *p* and *q* being necessarily such that the contemplation of both requires one to prefer *p* to *q,* then *p* will be intrinsically better than *q* whenever *p* and *q* occur. Similarly, Moore says with respect to intrinsic value:

> it is impossible for what is strictly one and the same thing to possess that kind of value at one time, or in one set of circumstances, and not to possess it at another; and equally impossible for it to possess it in one degree at one time, or in one set of circumstances, and to possess it in a different degree at another, or in a different set.[11]

Thus, according to Moore and Chisholm, if Smith's being pleased is intrinsically better than Brown's suffering, then the former will always be intrinsically better than the latter whenever the two occur. Of course, both Moore and Chisholm allow that it is possible that things have different instrumental values in different contexts and under different conditions. Suffering might have instrumental value if it deters from wrongdoing, but it can, in some circumstances, harden the heart, stiffen resolve, and fail to deter. The instrumental value of suffering is not essential to it, but if Brown's suffering is intrinsically bad, it is intrinsically bad whenever it occurs. In Chapter 3, we consider some views on intrinsic value that reject the thesis of universality.

What is it for a fact or state of affairs to be intrinsically good or bad? I suggest that we can explicate intrinsic goodness and badness, and other related value concepts, in terms of such concepts as "being intrinsically worthy of love" and "being intrinsically worthy of hate." But what is it for a state of affairs to be intrinsically worthy of love or hate? I propose to follow Chisholm in explicating such concepts in terms of the notion of "ethical requirement."[12] I shall also make use of the concepts of love and hate simpliciter, which Chisholm does not use, and that of preference simpliciter, which he does. Let us consider the following principles, where *p* and *q* stand for states of affairs and the attitudes of love, hate, and preference are taken to be attitudes simpliciter. Let us begin by accepting what Chisholm says about the concept of intrinsic preferability.

> (Pl) *p* is intrinsically worthy of preference to *q* if and only if *p* and *q* are necessarily such that, for any *x,* the contemplation of just *p* and *q* by *x* requires that *x* prefer *p* to *q*.

What is it for something to be intrinsically worthy of love or intrinsically worthy of hate? Let us say:

---

[11] G. E. Moore, "The Conception of Intrinsic Value," *Philosophical Studies* (London: Routledge & Kegan Paul, 1922), pp. 260-1; see also G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), p. 30.

[12] For a discussion of the concept of requirement, see Roderick M. Chisholm, "Practical Reason and the Logic of Requirement," in Stephan Körner, ed., *Practical Reason* (Oxford: Basil Blackwell, 1974), pp. 2-13; also in *Practical Reasoning*, edited by Joseph Raz (Oxford: Oxford University Press, 1978), pp. 118-127.

(P2) *p* is intrinsically worthy of love if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* requires that *x* love *p* and not hate *p*.

(P3) *p* is intrinsically worthy of hate if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* requires that *x* hate *p* and not love *p.*

There is, I think, a helpful analogy between a state of affairs being intrinsically worthy of love or hate and a proposition being an *axiom* or *self-evident.* What is it for something to be an axiom? Consider truths such as those expressed by "All men are men" and "If *A* is taller than *B,* then *B* is not taller than *A.*" Robert Audi writes, "Such truths are often called self-evident because they are obvious in themselves: if one comprehendingly considers them, one believes and knows them. One need not consult one's experience, nor even reflect on such propositions, to grasp – roughly to understand – them and thereby believe that they are true."[13] An axiom is a proposition that cannot be incorrectly believed and cannot be considered and accepted without its being evident to the person who accepts it. Thus, we may say that whoever accepts an axiom accepts correctly, "has taken the right stand," both from the standpoint of accepting what is true and from the standpoint of accepting what he is epistemically justified in believing. The analogy between an axiom and that which is intrinsically worthy of love consists in this: Whoever considers and accepts an axiom believes in a way that is *epistemically* required, and whoever considers and loves as such what is intrinsically worthy of love loves in a way that is *ethically* required. If one understands what it is for a proposition to be an axiom, and understands the concepts of ethical requirement and love simpliciter, one can understand what it is for something to be intrinsically good. I say more about the concept of an axiom and the analogy to what is intrinsically valuable in Chapter 8.

What is it for a state of affairs to be neutral? Roughly, a state of affairs is neutral just in case it is not intrinsically worthy of love and not intrinsically worthy of hate. I suggest, however, that we should distinguish between several different types of neutral states of affairs. Such distinctions are called for because the sorts of attitudes that are appropriate to various neutral states of affairs are not the same. We will use the following distinctions in discussing the value of things like pleasure, knowledge, and beauty. Let us consider the following two propositions:

(P4) *p* is *strictly* neutral if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* requires both that *x* not love *p* and that *x* not hate p.

---

[13] Robert Audi, *Belief, Justification, and Knowledge* (Belmont, Calif: Wadsworth Publishing Co., 1988), p. 51. See Chisholm: "*h* is an axiom = *Df. h* is necessarily such that (i) it is true and (ii) for every *S*, if *S* accepts *h,* then *h* is certain for *S.*" *Theory of Knowledge*, 2nd edition (Englewood Cliffs, N.J.: Prentice-Hall, 1978), p. 42.

(P5) *p* is *merely* neutral if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* does not require that *x* love *p* and does not require that *x* hate *p*.

(P5) is clearly a more latitudinarian notion than (P4). The difference between them is important. If a state of affairs is merely neutral, then it is possible for someone who has contemplated it to love it or hate it without violating any requirement. Similarly, to the extent that taking pleasure in a thing in and for itself appears to involve loving that thing simpliciter, it is possible for someone contemplating a merely neutral state of affairs to take intrinsic pleasure in it without violating any requirement. In contrast, anyone who takes pleasure in or loves simpliciter a strictly neutral state of affairs violates an ethical requirement in doing so. Consider the state of affairs *there being stones.* Suppose that someone contemplates just this state of affairs and loves or takes intrinsic pleasure in it. If *there being stones* is strictly neutral, then his loving or taking pleasure in it violates an ethical requirement. If, on the other hand, it is merely neutral, then his loving or taking pleasure in it does not. It seems that however odd or unusual it would be, one does not violate any ethical requirement simply by taking intrinsic pleasure in or loving simpliciter the existence of stones. Consequently, I believe that the state of affairs *there being stones* is merely rather than strictly neutral.

Are there any states of affairs that are strictly neutral? A likely candidate would be a state of affairs in which intrinsic goods and evils are "balanced off," for example, Smith's having five units of pleasure and Brown's having five units of pain. Given that this state of affairs involves as much good as evil and vice versa, it is appropriate that one not love or hate it simpliciter. Of course, it has "parts" that it is appropriate to love and to hate simpliciter.

In addition to the strictly and merely neutral, we may also distinguish between the *positively* and *negatively* neutral. Let us say:

(P6) *p* is *positively* neutral if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* requires that *x* not hate *p* and the contemplation of just *p* by *x* does not require that *x* love *p.*

(P7) *p* is *negatively* neutral if and only if *p* is necessarily such that, for any *x,* the contemplation of just *p* by *x* requires that *x* not love *p* and the contemplation of just *p* by *x* does not require that *x* hate *p.*

The importance of distinguishing the positively and negatively neutral can be brought out by reflecting on the absence of a good or evil. Consider the state of affairs *there being no one who is happy.* This state of affairs can be thought of as the absence of a good. Is this state of affairs intrinsically bad? Many writers have said that it is not. Oskar Kraus writes, "The nonexistence of a good is not an evil and the nonexistence of an evil is not a good; one can say only that the existence of

a good is *preferable* to its nonexistence, and conversely for evil."[14] The same view is taken by Chisholm and Ernest Sosa.[15] Suppose that these writers are correct and *there being no one who is happy* is *not* an intrinsically bad state of affairs. I suggest that the contemplation of this state of affairs requires that one not love it in and for itself. It is inappropriate to love in and for itself or to take intrinsic pleasure in the absence of a good. If this is right, then *there being no one who is happy* is not merely neutral. It also seems, however, that there is no requirement that one not hate the absence of a good. If this is so, then *there being no one who is happy* is not strictly neutral. I suggest that this state of affairs is negatively neutral. Similarly, consider the absence of an evil such as *there being no one in pain.* We may say that the contemplation of this state of affairs requires that no one hate it simpliciter, but the contemplation of it does not require that anyone love it simpliciter. It is an example of the positively neutral. We may say that the absence of an evil is a positive neutral and the absence of a good is a negative neutral.

We have distinguished four forms of neutral states of affairs. Shall we say that some forms of neutrality are intrinsically worthy of preference to other forms of neutrality, that, for example, anything that is positively neutral is intrinsically worthy of preference to anything that is negatively neutral? I see no reason to think so. I see no reason to hold that there being no one in pain is intrinsically preferable to there being no happy people. Holding that neither of these states of affairs is preferable to the other is quite compatible with there being a requirement not to hate the former and a requirement not to love the latter.

What is it for a fact to be intrinsically good or bad? Again, taking *p* and *q* to stand for states of affairs, let us say:

(P8) *p* is intrinsically good if and only if *p* obtains and *p* is intrinsically worthy of love.

(P9) *p* is intrinsically bad if and only if *p* obtains and *p* is intrinsically worthy of hate.

(P10) *p* is intrinsically neutral if and only if *p* obtains and *p* is not intrinsically worthy of love and not intrinsically worthy of hate.

(P11) *p* is intrinsically better than *q* if and only if *p* and *q* obtain and *p* is intrinsically worthy of preference to *q*.

If a fact is intrinsically good, then the contemplation of just that obtaining state of affairs requires that one not hate it in and for itself. Thus, if Smith's being pleased is intrinsically good and one contemplates that state of affairs, there is a requirement that one not hate it simpliciter. As Aristotle suggests, such a hatred would be

[14] Roderick Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986), p. 61.

[15] Roderick Chisholm and Ernest Sosa, "On the Logic of 'Intrinsically Better'," *American Philosophical Quarterly* 3 (July 1966), pp. 244-9.

ethically unfitting and a mark of the envious man. To be indifferent or not to love as such someone's being pleased is not precisely to be envious, but it is a mark of, at best, a small-minded niggardliness and, at worst, an ungenerous lack of goodwill to others. Similarly, the contemplation of something intrinsically bad requires that one not love it simpliciter. Thus, if someone's being in pain is intrinsically bad, then anyone who considers that state of affairs is required not to love it in itself. To love as such someone's being in pain is, according to Aristotle, spiteful and, according to Schopenhauer, diabolical. It is an ethically incorrect attitude. In contrast, it is a mark of the ethically good man to love as such what is intrinsically good and to hate as such what is intrinsically bad.

2. SOME OBJECTIONS TO THIS APPROACH

Let us consider four objections to the view that intrinsic value can be explicated in terms of required love, hate, and preference. The first objection is that in explicating the notion of intrinsic value in terms of an ethical requirement, we are confusing intrinsic value with *moral* value, that is, we are confusing intrinsic goodness with moral goodness and intrinsic badness with moral badness. But clearly, we are *not* confusing intrinsic goodness with moral goodness where the latter is construed as a certain kind of character or property of the person, the will, or certain actions. Of course, we are saying that the ethically or morally correct attitude toward what is intrinsically good is one of love rather than hate or indifference. Thus, if the suffering of an animal is intrinsically bad or the existence of a beautiful painting is intrinsically good, then for anyone who considers such things, the ethically correct attitude toward the former is one of hate and that toward the latter is one of love or favor. Presumably, a morally good person would not hate or be indifferent to what is worthy of love, or love or be indifferent to what is worthy of hate.

A second objection to this approach is the following: "Consider the following states of affairs: something's happening that is very likely to make 100 people happy (*p*) and something's happening that is very likely to make 100 people unhappy (*q*). It is false that *p* is intrinsically preferable to *q,* but the view previously described incorrectly implies that it is, for if one considers just *p* and *q,* then the contemplation of just *p* and *q* requires that one prefer *p* as such to *q*." The response to this objection is that the contemplation of just *p* and *q* does *not* require that one prefer *p as such* to *q*. *p* involves a state of affairs that is intrinsically worthy of love, namely, there being 100 happy people, and *q* involves a state of affairs that is intrinsically worthy of hate, namely, there being 100 unhappy people. But from the fact that there is a requirement to prefer as such there being 100 happy people to there being 100 unhappy people, it does not follow that there is a requirement to prefer *p* in itself to *q* in itself. We may also add that the contemplation of just *p* (unlike the contemplation of the different state of affairs, there being 100 happy people) does not require that one love *p* in and for itself.

A third objection urges that two things might be the same in intrinsic value, whereas the attitudes and feelings that are appropriate to one might be inappropriate

to the other. Consider the suffering of one's own child and the equal suffering of a child starving in Ethiopia. We may reasonably believe that the suffering of some child unknown and distant is every bit as bad as the suffering of our own child, but is it plausible to believe that the attitudes and feelings that are appropriate to the one state of affairs are also appropriate to the other? Brand Blanshard writes, "There are many feelings, felt with great intensity, that are obviously suitable in the case of our own child, which we could hardly be expected to show about one that was remote and all but unknown."[16] Toward the suffering of our own child we feel acute grief and distress, but only a certain melancholy and sadness about the suffering of faraway strangers, and we do so without any sense that the difference in our attitudes is inappropriate. But if the difference in our attitudes is appropriate, if it is fitting for us to have different attitudes toward these states of affairs, then how can we say that they have the same intrinsic value?

In response to this objection, one might urge that it is *not* fitting or appropriate for us to have different feelings toward our own child's misfortune and that of a total stranger. One might urge that the same acute grief or the same cool sadness, or perhaps something in between, is the fitting attitude to take toward both. Consider the advice of the Stoic Epictetus:

> when some other person's slave-boy breaks his drinking cup, you are instantly ready to say, "That's one of the things which happen." Rest assured, then, that when your own drinking-cup gets broken, you ought to behave in the same way that you do when the other man's cup is broken. Apply now the same principle to the matters of greater importance. Some other person's child or wife has died; no one but would say, "Such is the fate of man." Yet when a man's own child dies, immediately the cry is, "Alas! Woe is me!" But we ought to remember how we feel when we hear of the same misfortune befalling others.[17]

I cannot believe that this response is right. We often have at different times different emotional attitudes and feelings toward a single event and take these different attitudes and feelings to be appropriate. Different emotions and feelings toward the death of one's child or one's parents seem more appropriate to these events when they are recent than when they have receded into the past. It seems inappropriate to have precisely the same emotions and feelings about their death at different times, for there comes a point when it is appropriate not to feel acute grief and sadness, to let these feelings pass, and to feel in their place a sadness that is less intense. If different emotions and feelings can be appropriate to the same event when the intrinsic value of that event has not changed, there is no obvious reason why different emotions and feelings cannot be appropriate to different events of the same value.

---

[16] Brand Blanshard, *Reason and Goodness* (New York: Humanities Press, 1978), p. 287.

[17] Epictetus, *The Encheiridion*, in Oliver Johnson, *Ethics*, 6th edition (Fort Worth, Tex.: Holt, Rinehart, and Winston, 1989), p. 96.

Now even if we concede that it is more appropriate to have a more intense feeling of grief or sadness toward the suffering of one's own child than toward the suffering of a total stranger, this concession does not imply that we cannot explicate intrinsic value in terms of required love, hate, and preference. This is so simply because grief, sadness, and melancholy are not the same attitudes as love, hate, and preference simpliciter. It is not at all obvious that one's contemplation of just the states of affairs *my child's suffering* and *an unknown child's suffering* requires that one prefer the latter as such to the former.

It should be emphasized that claiming that there is no requirement that one prefer in itself the latter to the former *is* compatible with holding that there is a requirement to alleviate the one rather than the other. It has already been noted that preference simpliciter is not the same thing as choosing, and which case of suffering one is required to alleviate depends on considerations other than the contemplation of just those states of affairs. These other considerations might include the nearness of the children, the probability of success in alleviating their pain, the costs of doing so, and even considerations of loyalty to one's children. But even if, in light of these other factors, there is a requirement to choose to alleviate the suffering of one's own child, it does not follow that one is required to prefer simpliciter the suffering of the unknown child.

Let us consider a fourth objection to the view that intrinsic value can be explicated in terms of required love, hate, and preference. According to this objection, if something merits favor, it can only be because or in virtue of the fact that it is good, and if it is correct or fitting to prefer one thing as such to another, it can only be because the former is better in it itself than the other. Blanshard says, "If saintliness and generosity are such as to merit favoring, it must be because there is something in them that goes beyond their 'factual characteristics' and equally goes beyond a mere blank cheque on our favor. What is this? I think we must answer, a goodness that they have already."[18] According to this objection, if it is fitting to favor something only because it is good already, if something's being good is the ground or reason why it is fitting to favor that thing, then the goodness of that thing cannot be identical with its being fitting to favor it and its being good cannot be defined in terms of its being fitting to favor it.

Strictly speaking, this is not an objection to the view presented earlier. Rather, it is an objection to the view that we can *eliminate* the concept of goodness in favor of the concepts of correct or required emotion. The objection insists that we cannot eliminate intrinsic goodness in this way because it is the goodness of a thing that explains why it is a required object of love. Of course, this is not an objection to the view I have offered, since I have only tried to explicate intrinsic goodness, not eliminate it. Still, this objection is not fatal even to those who do propose to eliminate goodness in favor of required attitudes. The mistake that underlies this objection is the assumption that it can only be the value of a thing that makes it required or appropriate to favor it. Why must there be something else in addition to

[18] Blanshard, *Reason and Goodness*, p. 287.

the "factual characteristics" of saintliness, generosity, wisdom, or pleasure that makes them worthy of favor, that makes it appropriate to look favorably on their instances? Surely it is plausible to think that there are other sorts of evaluative and normative features of things that they have in virtue of their factual or nonevaluative characteristics. The goodness of an apple would seem to depend on its being sweet, juicy, and so on. The beauty of a painting would seem to depend on such factual characteristics as its colors and their arrangement. The rightness of an action may depend on its factual characteristics, such as whether it produces a greater balance of pleasure over pain than any alternative, whether it is the keeping of a promise or an instance of gratitude or loyalty.There is, therefore, no obvious reason why the fittingness of an emotional attitude should be thought to depend on something other than the nonevaluative nature of its object. If so, there is no obvious reason to think that the fittingness of an attitude must depend on an additional distinct feature or characteristic of value. Moreover, it would seem hard to deny that what has value has it because or in virtue of its having some factual, nonevaluative characteristic or because of the obtaining of some nonevaluative state of affairs. But if value is itself dependent on the factual and the nonevaluative, then it is strange to argue that meriting favor or disfavor cannot be similarly dependent. Why should there be any difference between the two? As far as I can see, there simply isn't any.

CHAPTER 5

J. DANCY

SHOULD WE PASS THE BUCK?

My topic is the relation between the right and the good. I introduce it by relating some aspects of the debate between various British intuitionists in the first half of the present century.

1.

In *Principia Ethica* (1903) G. E. Moore claimed that to be right is to be productive of the greatest good. He wrote 'This use of "right", as denoting what is good as a means, whether or not it be also good as an end, is indeed the use to which I shall confine the word' (p. 18). By the time he wrote his *Ethics* (1912, e.g. p. 6) he seems to have weakened his position, and offers conduciveness to the good not as a definition of 'right' but as an account of the one and only property that makes acts right. Even if it be the only right-making property, conduciveness to the good will not be identical with the rightness that it makes.

One might ask why Moore changed his view, and an obvious answer is that he came to see that the notorious Open Question Argument, by which he strove to establish that goodness is not identical with any good-making feature, can be used equally well to show the same thing of rightness. If it is an open question whether goodness is conduciveness to happiness, it is equally an open question whether rightness is conduciveness to goodness. And if, as Moore claimed in the first case, its being an open question shows that the answer to it is no, the same applies in the second case.

W D. Ross details all these matters with further references, and with his customary clarity, in the early pages of his *The Right and the Good* (1930, pp. 8-11). Though he argues that Moore's second view is a vast improvement on his first, Ross's own position is quite different. Rightness and goodness are utterly distinct; indeed, no one thing can be both right and good. Goodness, for Ross, is a property of motives and outcomes, and rightness is a property of acts. An act can be intrinsically right or wrong, but never intrinsically good or bad. Acts can be instrumentally good, or conducive to good; but Ross announces, surely correctly, that instrumental value is not a form of value at all. So it turns out that acts can have no value at all. Motives, by contrast, can be intrinsically good or bad but never right or wrong. On this picture the very idea that one might define the right in terms of

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 33-44.
© 2005 *Springer. Printed in the Netherlands.*

the good is quite peculiar. Also peculiar is the idea that the only way that an action could get to be right is by having the best consequences.

H. W. B. Joseph's response to this in his *Some Problems in Ethics* (1931) was that Ross's position was absurd. 'Why', he asked, 'ought I to do that, the doing which has no value (though my being moved to do it by the consciousness that I ought, has), and which being done causes nothing to be which has value? Is not duty in such a case irrational?' (p. 26). Rightness, for Joseph, must be in some way dependent on goodness. He pursued this idea by claiming that the word 'right' is ambiguous. In one sense it means 'obligatory', and Joseph writes (p. 61) that 'obligatoriness is not a character of actions. There is no ought-to-be-done-ness, or ought-to-be-forborne-ness. To say that an act is obligatory means that the doing it is obligatory on me.' In the other sense, 'rightness is a form of goodness, to the realising of which the actions belong; and it is the thought of goodness which moves us when we do an action from a sense of obligation' (p. 104). In this way Joseph rejects Ross's claim that rightness and goodness are utterly distinct.

In his second book *The Foundation of Ethics* (1939) Ross's position becomes more complex. He has already argued that there are two uses of 'good', attributive and predicative. The attributive use is at issue when we speak of a good liar or a good knife. The predicative use is the one that is of importance for ethics, and it is found when we speak of a good man, or claim that virtue, knowledge and pleasure are good. Ross claims that when we say that the pleasure of others is good, we mean that it is a proper object of satisfaction. This is a 'definition' of this use of 'good'. When we speak of a man, or of a motive, as being morally good, however, we mean something else, something that cannot be defined but only paraphrased (p. 283). The paraphrase is that the good, in this use of 'good', is a proper object of approval, worthy of approval or admiration.

Why is this not a definition? Because Ross is still sticking to his original view that goodness, in *this* sense, is an intrinsic property of objects, not a relation. If being good in this sense were being worthy of approval or admiration, it would be a relation. But it is not; this sort of goodness is the property that in approving or admiring we take the object approved or admired to have. For to approve is to think good, and 'admiration is not a mere emotion; it is an emotion accompanied by the thought that that which is admired is good' (pp. 278-9).

It is worth pausing to note what Ross means by a relation here. It is not what we would ordinarily mean, because we would ordinarily think that for a relation to obtain, there must be at least two relata and both must, in the relevant sense, exist. But in suggesting (even if only to reject the idea) that goodness might be a relation, Ross is clearly not thinking of relations in this way. For something can be worthy of approval even if no approval and no approver is forthcoming.[1] Further, though Ross thinks that goodness in this sense is not a relation, that is not his reason for rejecting the claim that goodness is identical with being worthy of admiration and approval. Ross's real point is that the goodness that we take the object to have cannot be identical with its being worthy of our so taking it, because it must be that in the object that makes our so taking it an appropriate or fitting response.

[1] It is just possible that Ross thinks of the relevant relation as 'being worthy of *our* approval', which would bring him back into line on this point.

It would have been possible to avoid this result if we had been more catholic in our choice of attitudes or responses that the good action is worthy to elicit. In his *The Definition of Good* (1947), A. C. Ewing defines the good as that which ought to be the object of a pro-attitude (pp. 148-9). He attributes the term 'pro-attitude' to Ross, and continues '"pro-attitude" is intended to cover any favourable attitude to something. It covers, for instance, choice, desire, liking, pursuit, approval, admiration. ... When something is intrinsically good, it is (other things being equal) something that on its own account we ought to welcome, rejoice in if it exists, seek to produce if it does not exist. We ought to approve its attainment, count its loss a deprivation, hope for and not dread its coming if this is likely, avoid what hinders its production, etc.' (p. 149).

By stressing the broad variety of attitudes that may count as pro-attitudes, Ewing seems to avoid Ross's only argument that we cannot define the good (in the relevant predicative sense) as what is worthy of a certain response. Ewing seems, that is, to be in a position to say that goodness is not a distinct evaluative and intrinsic property in objects, one whose presence we can discern and to which we do or at least should respond with approval and admiration. The goodness of the object just is the relational fact that we should respond to it with approval, admiration or other pro-attitude. The evaluative 'good' has been defined in terms of the deontic 'should'. And with this result, the intuitionists reversed Moore's position, the position with which I started this brief history. Moore defined the right, that which we ought to do or should do, in terms of the good. Ewing defined the good in terms of how we should respond.

This is the end of my brief historical introduction. At the point we have reached (the late 1940s) the intuitionists' broadly cognitivist approach to ethics was eclipsed by the onrush of non-cognitivism. Fifty years passed before these sorts of issues could again be debated with any sense of seriousness.

Fifty years on, Thomas Scanlon writes as follows in his *What We Owe to Each Other* (1999, pp. 95-7):

> To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it. ... To say that something is valu*able* is to say that others also have reason to value it, as you do. ... this account … takes goodness and value to be ... the purely formal, higher-order properties of having some lower-order properties that provide reasons of the relevant kind. [It holds] that it is not goodness or value itself that provides reasons but rather other properties that do so. For this reason I call it a buck-passing account.

My question is whether we should in this way seek to pass the buck. Should we take it that to be valuable is to have features that give us reasons, in the way that Scanlon suggests?

2.

I start with some comments on the buck-passing view as expressed above by Scanlon. First, I think there must be a slip in the way that Scanlon tries to capture the buck-passing account of taking something to be valuable. It cannot be right to

say that to take something to be valuable is to take it that others also have reason to value it, as you do. For in valuing it we do not take ourselves to have reasons to value it; at least, not on the buck-passing view. On that view, to value it is to take oneself to have reasons of certain other sorts. The quotation I gave above does not reveal this, but Scanlon does detail these sorts of reason. He talks of reasons for admiring, respecting, preserving and protecting; of reasons to be guided by the goals or standards that the value involves; of reasons for promoting; and of reasons to act in certain ways. To value something is to take oneself to have reasons of these sorts, reasons that are given one by features of the object (or perhaps more broadly of the situation) which act as ground for the reasons and, in doing so, as ground for the value. But when we value the object we are not taking ourselves to have reason to value it, exactly. So when we come to think of the valuable, rather than just of the valued, we should surely be thinking of others as having the same reasons as we do, reasons to admire, protect, promote, etc.

With that slip corrected, Scanlon's view is very similar to Ewing's. Scanlon's is expressed in terms of reasons while Ewing spoke of how we ought to respond. But a definition of value in terms of reasons is very similar to one in terms of oughts, especially when we notice that Ewing was speaking of *prima facie* oughts, and Scanlon is speaking of what one might call contributory reasons. Crucially, the notion of what one ought to do is a deontic notion, as I take that of a reason to be, and the notion of a value is evaluative; Ewing and Scanlon combine to give a deontic definition of the evaluative, and this is what is at once attractive and worrying about their position.

What I mean by this is that we should not forget the position they are trying to undermine. We might think that there are just two normative concepts. The first family is the evaluative family. Here we are dealing with the good and the bad, the noble, the fine and the evil. Second is the deontic family. Here we are dealing with the right and the wrong, shoulds, oughts and musts, obligations, requirements and prohibitions. Ross took the view that these two families are utterly different. His picture of the normative realm is therefore far more complex than is that of Ewing or of Scanlon – or of Moore, of course, at the other end. Complexity has its advantages and its disadvantages, as we will see. But we may be permitted to hope that the Ewing/Scanlon view, the buck-passing view, has not only the advantages of simplicity. If we are to accept it, there must be more to be said in its favour than just that.

And of course there is. At least, there is if Scanlon is to be believed. He argues for the buck-passing view by arguing against two alternative positions, supposing that if they fail the buck-passing view will emerge as the sole remaining contender. Unfortunately, I will now suggest, he does not consider every available alternative. So his argument by remainder ends up looking a bit weak.

The opposing views, according to Scanlon, are (1) 'the teleological view' and (2) 'the view that value is a property the presence of which grounds or explains reasons'. Interestingly, these two views look very much like the earlier and the later views of Moore. In Moore's terms, the 'teleological view' is the view that rightness is to be defined in terms of value rather than the other way around. The second view is the view that value is the right-making property, though not identical with

rightness. Expressed in terms of the relation between values and reasons, rather than between value and rightness, the teleological view will be that reasons are to be defined in terms of some relation to value, and the second view will be that goodness is a (or the) reason-giving property. Now I accept that both of these views are wrong. We abandon the teleological view because we suppose that we cannot define reasons in terms of values, whatever the proper account of the relation between reasons and values turn out to be. We abandon the second view because we recognise immediately that the badness of a toothache, for example, does not add a further reason to the reasons for going to the dentist that are already given by the nature of the toothache, its painfulness. As we might put it, the badness of the toothache exists in virtue of certain features, features which give us reason to act in certain ways; the badness of the ache adds nothing to the reasons given us by those lower-level features. In short, value adds no reasons to those generated by the ground for that value.

Does it follow from this that to be of value is just to have reason-giving features? I do not see that it does. Two further views remain unrefuted. The first is Ross's view, expressed in terms of the relation between reasons and value rather than in terms of the relation between rightness and goodness. Values and reasons are just utterly different, and neither can be defined in terms of the other. This was clear, for Ross, because the things that are duties have no value at all, and the things that have value cannot be duties. There is indeed a link between values and reasons, since where there is intrinsic value we *always* have a reason. Ross, after all, thought that 'one of our main duties is to produce as much that is good as possible' (1939, p. 257). The prospect of value always gives us a reason, then, and this is something that needs to be explained if values are as different from reasons as Ross thought they were from duties. But Ross incurs this explanatory debt by adopting a picture for which he thought there were perfectly good independent reasons.

The second view that remains is one that, without committing itself to any vast difference between values and reasons, supposes none the less that the evaluative is distinct from the deontic, even if both sorts of properties result from the same ground. If they result from the same ground, one and the same object can have both evaluative and deontic properties; it can be good, and can be something that we have sufficient reason to do. This final view incurs the same explanatory debt as Ross's did, if it is combined with the claim that a ground for value is always (or even therefore) a ground for reasons as well.

Here then is a map of the five possibilities I have mentioned, where *r* = reasons, *v* = value and *g* = ground, and the relation '→' is the grounding relation:

1. Early Moore (teleological view): $g \rightarrow v = r$ (reasons defined in terms of value)
2. Later Moore: $g \rightarrow v \rightarrow r$ (reasons grounded in but not defined in terms of value; value adds to the reasons given by the ground)
3. Ross: $g_1 \rightarrow v$, $g_2 \rightarrow r$ (i.e. the ground for reasons differs from the ground for value)
4. Buck-passing View: $g \rightarrow r$
   $\uparrow$
   v (i.e. being of value is having features that ground reasons)
5. Last View: $g \rightarrow v$, $g \rightarrow r$ (reasons and values are distinct but may have the same ground).

My only point here is that one does not establish the buck-passing view by refuting the two views of Moore. If that is right, and fair criticism of Scanlon, the main argument in favour of the buck-passing view will be its theoretical neatness. This neatness largely consists in the fact that it does not require us to say why it is that wherever there is value in the offing, there are at least some reasons. Ross's view is vulnerable to this requirement, as is the Last View.

At this stage, I need to introduce a different topic, attitudes to which will help us to decide between the buck-passing view and its alternatives. This topic is the relation between ground, rightness and reasons.

Remember that we agreed with Scanlon that the badness of the pain cannot add to the reasons generated by the features that make the pain bad (i.e. effectively, by the way it feels). Now are we to say the same sort of thing about rightness and wrongness? Are we to suppose, that is, that though goodness and badness cannot add to the reasons coming up from below, rightness and wrongness can? Should we suppose that the thin properties (or relations) are in general incapable of adding to the reasons thrown up from below? It is perhaps a bit surprising that Scanlon's answer to this question is no. According to Scanlon, the wrongness of an action can add to the reasons not to do it, though the badness of the action cannot. There are the reasons given us by the features that make it wrong, and then there are further reasons given us by the fact that it is wrong, or by its wrongness.

Even if, considering the matter simply in structural terms, we would probably expect all the thin properties to behave as reason-givers in much the same sort of way, that expectation could be overturned by a closer investigation of what actually happens. Derek Parfit considers the relation between two claims:

(1) This act violates standards of conduct that we all have very strong reasons to regard as important and to follow.

(2) This act gives the agent reasons to feel guilty and gives others reasons for indignation and resentment.

Suppose now that (1) specifies what it is for an act to be wrong. Parfit continues 'if our violating such standards would give us a reason to feel guilty, and would give

others reasons to be indignant or resentful towards us, these facts do seem to give us further reasons not to act in this way'.[1]

If Parfit is right, the picture of the relation between ground, wrongness and reasons is like this:

$$g \rightarrow w \rightarrow r$$

While if the alternative view is right, it is more like:

$$g \rightarrow r = w$$

This is because those who think that the wrongness of an action cannot add to the reasons against doing it that are thrown up from below, are motivated by a certain conception of what wrongness and rightness are. We might call this the 'verdictive conception'.[2] In deciding whether an action is right, we are trying to determine how the balance of reasons lies. Our conclusion may be that there is more reason (or more reason of a certain sort, perhaps) to do it than not to do it, and we express this by saying that it is therefore the right thing to do. The rightness-judgement is verdictive; it expresses our verdict on the question how the reasons lie. It is incoherent, in this light, to suppose that the rightness can *add* to the reasons on which judgement is passed, thus, as one might say, increasing the sense in which, or the degree to which, it is true. And the same is true of wrongness. So when I expressed the 'alternative view' above as I did, I meant by this to show how that view understands wrongness-judgements as expressing, or capturing, how things are at the level of reasons. And I supposed that if this is correct, the wrongness of the action cannot itself add to the reasons that it captures.

We might think of this as a buck-passing view about rightness and wrongness. Scanlon wrote that his view 'is called the "buck-passing view" because it takes the normative force of a claim that something is good or valuable to be inherited from that of the reasons which it asserts to be present'. To claim that something is good is to claim, according to the buck-passing view, that it has features that give us (*pro tanto*) reasons to take certain attitudes to it. Similarly, we might say, to claim that an action is right is to claim that it has features that give us overall reason to do it. This is a buck-passing view because it holds that the normative force of a claim that something is right or wrong is inherited from that of the reasons which it asserts to be present. So for Scanlon to adopt what I called the 'alternative view' about the relations between rightness and reasons would, one would have thought, have been for him to be a consistent buck-passer.

Think now about the arguments that Parfit produced in support of the non-buck-passing conception of rightness and wrongness. They amounted to the claim that if the wrongness of our action would give us a reason to feel guilty, and would give others reasons to be indignant or resentful towards us, these facts give us further reasons not to act in this way.

---

[1] This quotation is taken by permission from a draft of Parfit's forthcoming *Rediscovering Reasons.*

[2] I take this name from Stratton-Lake (2000), chapter 3.

Should these arguments persuade us not to be consistent buck-passers? I doubt it. The same arguments seem to apply to badness. Suppose that my acts are bad as well as wrong. I would say that the badness of my behaviour gives me reason to feel ashamed, and gives others reasons to take certain attitudes towards me. Won't those facts then give me further reasons not to behave in those ways, if this is what happens in the case of wrongness as Parfit suggests?

Alternatively, we could look again at the suggestion that if the wrongness of our action would give us a reason to feel guilty, and would give others reasons to be indignant or resentful towards us, these facts give us further reasons not to act in this way. It would be possible to hold that the wrongness of our action does not give us reason to feel guilty. What it does is to show that if our action is wrong, we have reasons to feel guilty, and others have reason to feel indignant or resentful towards us. It does not show that the wrongness of our action adds to those reasons.

Be this as it may, we have now come across three possible overall views. The first is Scanlon's and Parfit's. They are buck-passers about the good, but not about the right. I have been suggesting the possibility of a more consistent buck-passing view, which passes both bucks. But I have also aired the possibility of a view that passes the rightness buck but not the goodness buck. This is because I have yet to be convinced of the buck-passing view for goodness. In the final section of this paper I try to say why.

3.

The points I raise here are not supposed to constitute a refutation of the buck-passing view for goodness. (I will call this just the 'buck-passing view' from now on, forgetting any problems there may be about passing the rightness buck.) They are supposed only to raise doubts. I try to raise doubts by pointing to significant areas of debate in meta-ethics which the buck-passing view would automatically resolve. My suggestion is that this is probably not the right way to resolve these issues, and that they would be better resolved by argument than by a sort of peremptory definition. Of course, if the buck-passing view were true, the issues would indeed be resolved, and that would be the end of the matter. But remember that the only effective argument we have seen for the buck-passing view is that it offers a pleasing theoretical neatness. I like neatness, but I do not like to see apparently significant issues resolved by a definition whose only recommendation is that of neatness.

The first point here involves the contrast between consequentialism and deontology. The buck-passing view threatens to resolve this debate in favour of consequentialism. Deontologists have suggested in one way or another that there are duties, and so reasons, that are not value-involving. An action can be one's duty even though doing it has no value and its being done generates nothing of value. Standard examples here are of trivial duties. Suppose that I promise my children that I will tie my right shoelaces before my left shoelaces on alternate days of the week if they will do their homework without fuss. One can imagine arguing that though I ought to tie my right shoelaces before my left shoelaces today, since I did the opposite yesterday, my doing so has no value of any form. The buck-passing

view rules this out in advance. To have value is to have reason-giving features, we are told, and since this is an identity statement it goes both ways. So to have reason-giving features is to be of value. So the deontological view expressed above is ruled out in advance of any significant debate.[3]

We might try to recover the issue we thought we had, by recasting the debate between consequentialism and deontology as a dispute about whether actions are made right solely by the value of their consequences, or whether they cannot also be made right by their intrinsic value.[4] This debate is still alive, even if the buck-passing view is true, for on both sides of the issue there is the sort of relation between reasons and value that the buck-passing view can allow. But there would still be one position on this issue that's ruled out in advance, namely that of Ross (and also that of Prichard, one might say). For Ross would not accept that actions that are right are made right by their intrinsic value, or by the features that give them that value. As I briefly pointed out at the beginning of this paper, his view was that no duty has value of any sort, and that nothing that has value is a duty. And he had independent philosophical reasons for the two parts of this view. Can it be right to say that we know his view to be false because of the attractions of the buck-passing view, when those only amount to theoretical neatness?

There is a further point of the same sort, which concerns the notion of agent-relative value. There were in fact two reasons for the initial deontological view that reasons are not essentially related to any sort of value. The first I alluded to above, namely the possibility of valueless but still right acts. The second is a worry about maximising. Suppose that we allow that every right action has some sort of value, just in virtue of being right. It should then turn out that we have reason not to do the action ourselves if we can thereby enable two other agents to do actions of similar value. And this is at odds with the basic deontological picture of duties. This picture is built on a distinction between two sorts of rule. To take an example that I think I owe to Philippa Foot: there is the rule 'do not shout' and there is the rule 'see to it that as little shouting as possible takes place'. The first of these is a deontological rule; it is addressed to the agent, saying, as it were 'don't shout – this means you' or 'don't be a shouter'; and we break that rule if we shout. The second rule is quite different. We do not necessarily break this rule by shouting; indeed, we would possibly only be able to keep the rule by shouting, as when we need to shout in order to shut everyone else up. In this sense, deontological rules are not maximising rules. You should not shout even if, by shouting, you can minimise the incidence of shouting. A maximising rule is not the sort of rule that deontologists take themselves to be talking about. But if we introduce a link between reasons and

[3] Derek Parfit has suggested to me that one could avoid the thought that the buck-passing view, being an identity claim, goes both ways by distinguishing two sorts of reasons – or, probably better, by distinguishing two sorts of things that reasons can be reasons to 'do'. I suggested above that the buck-passing view is pretty catholic about the sorts of reasons it is talking about, talking of reasons for admiring, respecting, preserving and protecting; of reasons to be guided by the goals or standards that the value involves; or reasons for promoting; and of reasons to act in certain ways. If some of these sorts of reason are involved in value, as we might put it, and others are not, the identity between having reason-giving features and being valuable will fail. I agree that this is so, but my present view is that there is no effective way of carving the reasons up appropriately.

[4] Philip Stratton-Lake suggested this to me.

values, we undermine this deontological picture of rules (obviously, and especially, of moral rules). For we reintroduce the possibility that the value at issue in keeping the rule is best served by breaking it.

There is, however, a possible way out of this difficulty. It involves the introduction of agent-relative value. If the sort of value that doing one's duty has is agent-relative value, it might be possible to shatter the maximising picture even though one has retained a general connection between reasons and value of the buck-passing sort. Now I don't want here to go into the details of how this might be done, if indeed it can be done at all. My point at this juncture is going to be merely that, if we are to try to prevent the adoption of the buck-passing view from undermining a significant aspect of deontology by introducing a conception of agent-relative value, this is a considerable theoretical cost, and it is a cost, once again, that we have committed ourselves to paying just for the sake of theoretical neatness. What is more, many people doubt the coherence of the notion of agent-relative value in the first place. If the buck-passing view can only be sustained by introducing a piece of dubious philosophy, it is looking much less attractive.

I turn now to a completely different source of disquiet about the buck-passing view. There is a history to this too. For Ross held that goodness is an intrinsic property, while rightness is a relation. Rightness is the relation of being fitting to the situation (1939, p. 52), while goodness is a property of motives and that is not a relation at all. Moral goodness, in particular, is a monadic property of 'acts of will, desires, and emotions, and finally relatively permanent modifications of character even when these are not being exercised' (1939, p. 292). Had such goodness been able to be defined as whatever is worthy of admiration and approval, it might have been seen as a relation (though, as discussed earlier, this might not be quite what we would now mean by a 'relation'). But we saw that Ross refused to allow this 'definition', though he did not reject it as a paraphrase.

I don't mean here to appeal to the detailed argumentation by which Ross defends his position. I don't even mean to defend his view that goodness, at least moral goodness, is not a relation at all. The real point, I think, is one about the polyadicity of rightness and of goodness – and this is a point that translates into thoughts about the polyadicity of reasons. Let us allow, without asking why for the moment, that rightness is a many-place relation. The point will then be that even if goodness is also a many-place relation, it has fewer places than rightness does and fewer than reasons do. Now if this is true, it cannot be correct to define goodness as the presence of reason-giving features. For the presence of reason-giving features will have more places in it, so to speak, than the goodness has.

The reason for supposing that goodness is less polyadic than reasons is that reasons belong to, are for, individuals. There are no reasons hanging around waiting for someone to have them. If the situation generates a reason for action, it must allot that reason to someone. (I don't mean to suggest that this is always or often difficult.) But goodness is not like this. Something can be good or bad without specification of an agent. The desolation or destitution of someone is bad even if there is nobody around to do anything about it, nobody who has any opportunity to do anything about it, and so nobody who can be said to have a reason to do something about it. Someone's destitution, then, has features that *would* ground

reasons for any agent suitably situated, but it does not follow that those features already ground reasons. And if it does not, we can be sure that to have value or disvalue is not itself to have reason-giving features.

One might reply that there are surely reasons for the destitute person, created by his destitution. In that case, a better example might be the lonely and sudden death of someone without friends or relatives, far from any possible help. We could say that this is bad, even if there is nobody who has reasons to grieve or indeed reasons of any other sort.

Derek Parfit suggested to me in conversation that there may be a weaker form of the buck-passing view that is consistent with the original motivation but which is less vulnerable to this sort of worry about polyadicity. This weaker form understands having value, not as having reason-giving features, but as having features that are potentially reason-giving. To have potentially reason-giving features is a less polyadic matter than to have actually reason-giving features, since something could be of the former sort without our needing to specify any particular individual for whom the reasons are reasons (since the reasons don't yet exist). This manoeuvre seems sound, so far as that goes, but it raises worries of another sort. For it seems far too easy to have features that are potentially reason-giving – that would, in certain circumstances, give us reasons. Something that has no value at all might well have features that would, in certain circumstances, ground reasons. So this weakening of the buck-passing view seems to me to enfeeble it.

The failure of these replies seems to me to establish that goodness and reasons do not have the same degree of polyadicity. Is that result consistent with the buck-passing view? I doubt it. To have features that ground reasons, where reasons are polyadic to degree *n*, is, I suggest, necessarily itself a property that is polyadic to degree *n*. Just having those reason-grounding features is, of course, not necessarily polyadic at all. But that is irrelevant. Being in pain may be the reason-grounding property, without itself being other than monadic. But that this feature is reason-giving must itself share the polyadicity of the reasons given, since it is a feature that can only be instantiated when the various empty spaces in the specification of the reason have all been filled up.

4.

My overall conclusion, then, is, that the buck-passing view needs more defence than it has so far received. The buck-passing view I am talking of is, of course, the view that to be good, or valuable, is to have features that give us reasons of certain sorts. The discussion above tended to favour the other buck-passing view, the view that for an action to be right is for it to have features that give us overall reasons (perhaps of a certain sort) in favour of doing it. If we yield to this pressure, neither we nor Scanlon will be consistent buck-passers. We will pass only the rightness buck, and he passes only the goodness buck. And a challenge might here be mounted in favour of consistent buck-passing.

In reply to that challenge, the appropriate response is to point out that rightness and reasons are concepts of the same normative family, the deontic family. It is not, therefore, so surprising if there are the sorts of links between them that are involved

if we pass the rightness buck. There would be no worries about polyadicity, for instance, for all the deontic concepts seem to be polyadic to the same degree. Equally, the deontologists may feel that goodness does not always generate reasons, or, more probably, that reasons do not always pass via goodness. The maximising worries that support this sort of thought have nothing to say against passing the rightness buck. So I think that, although it would be prettier if we could pass both bucks, and that there is an initial attraction in the cry 'both bucks or neither!', the terrain is not flat enough to allow us to do this.

Refusing to pass the goodness buck does nothing to reinstate something that we all think mistaken, the thought that the value of an object adds to the reasons that come from below, the reasons given us by the features that generate the value. This thought is a mistake, but that does nothing to support Scanlon's buck-passing view. Nor can we defend that view by asking 'how can there be value without reasons?' (the opposite of Joseph's question how there can be reasons without value). This question is irrelevant. The question before us has been whether to be valuable is to have reason-giving properties, not whether all valuable things are ones that have features that give us reasons to treat them in one way rather than another. The latter is not disputed. But to allow that anything with value will have reason-giving features is not to accept the buck-passing view. These things may go regularly, constantly or even invariably together without being identical.

Suppose, however, that they do go invariably together. Or suppose that to be valuable is to have what we might call 'potentially reason-giving features', ones set up to give reasons for any suitable agent, should there be one. These thoughts create a link between values, between being valuable, that is, and reasons. And we still need to find a way of explaining this link. My final remark is the admission that on this score the buck-passing view has an advantage, though not, I think, a final one. All reductive views, after all, give peremptory answers to questions about interconnections between the 'two' features that they reduce to one, and those peremptory answers are often, as here, unsatisfying.[5]

REFERENCES

Ewing, A. C. (1947) *The Definition of Good* (London: Macmillan).
Joseph, H. W B. (1931) *Some Problems in Ethics* (Oxford: Clarendon Press).
Moore, G. E. (1903) *Principia Ethica* (Cambridge University Press).
Moore, G. E. (1912) *Ethics* (Cambridge University Press).
Ross, Sir W. D. (1930) *The Right and the Good* (Oxford: Clarendon Press).
Ross, Sir W. D. (1939) *Foundations of Ethics* (Oxford: Clarendon Press).
Scanlon, T. M. (1999) *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press).
Stratton-Lake, P. (2000) *Kant, Duty and Moral Worth* (London: Routledge).

[5] Many thanks to Derek Parfit and to Philip Stratton-Lake for many discussions of these topics and helpful criticism and suggestions.

CHAPTER 6

F. FELDMAN

HYPERVENTILATING ABOUT INTRINSIC VALUE*

1. HYPERBOLIC REMARKS ABOUT INTRINSIC GOODNESS

The friends of intrinsic value (and here I include myself) often lapse into poetry when they try to describe the object of their common interest. They speak in metaphor, analogy, and hyperbole. Plato, for example, gave analogies, saying that The Good is in some way like the sun.[1] He suggested that each is a source of immense value. And just as the sun is too blinding to observe directly with the naked eye, so the form of The Good is too dazzling to contemplate directly with the naked mind.

In a particularly high-flying passage, he compares a philosopher who has grasped the concept of goodness with a cave-dweller who has emerged from his subterranean cavern into the blinding light of the sun. He says:

> In the world of knowledge, the last thing to be perceived and only with great difficulty is the essential Form of Goodness. Once it is perceived, the conclusion must follow that, for all things, this is the cause of whatever is right and good; in the visible world it gives birth to light and to the lord of light, while it is itself sovereign in the intelligible world and the parent of intelligence and truth. Without having had a vision of this Form no one can act with wisdom, either in his own life or in matters of state.[2]

Those who perceive this blinding form of goodness will return to their cavern dazed and confused. Their former compatriots will think them ridiculous, and may even try to kill them.

Immanuel Kant likewise drew comparisons. In describing a thing he took to be good in some outstandingly fundamental way, he tried to make clear that this thing does not have its value because of its capacity to produce good results. For even if "by the niggardly provision of a stepmotherly nature" it were to have no extrinsic value at all:

---

* Paper presented at the 21st Annual Greensboro Symposium in Philosophy, April 5, 1998. I thank Alison McIntyre, Michael Zimmerman, Earl Conee, and Owen McLeod for helpful comments.

[1] Plato, *The Republic*, translated and with an introduction and notes by Francis MacDonald Cornford (New York and London: Oxford University Press, 1958), p. 231.

[2] Ibid.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 45-58.
© 2005 *Springer. Printed in the Netherlands.*

> ...it would still sparkle like a jewel in its own right, as something that had its full worth in itself... Its usefulness would be only its setting, as it were, so as to enable us to handle it more conveniently in commerce or to attract the attention of those who are not yet connoisseurs, but not to recommend it to those who are experts or to determine its worth.[3]

G. E. Moore himself seemed to have trouble finding clear, literal words to describe this object. In one place, while struggling to express himself, he said this: "If I am asked 'What is good?' my answer is that good is good, and that is the end of the matter. Or if I am asked 'How is good to be defined?' my answer is that it cannot be defined, and that is all I have to say about it."[4]

It seems to me that we simply must do better than this. Metaphors and analogies and enforced silences will not help us to understand each other or the object of our common interest. I fear that by speaking in these dark ways, we give aid and comfort to those who are no friends of intrinsic goodness. They are inclined to think that we don't know what we are talking about (or maybe that we are talking about nothing at all).

It also seems to me that insofar as there is confusion about intrinsic goodness, there is even greater confusion about the other concepts of axiology and perhaps also those of normative ethics. For it is reasonable to suppose that other value concepts (extrinsic goodness, signatory goodness, etc.) are to be defined by appeal to the concept of intrinsic goodness. It is also at least somewhat reasonable to suppose that the concept of intrinsic goodness plays some role in the explanation of such things as moral rightness of actions, and virtuosity of character.

Thus, there is reason for us to insist upon clarity and precision in our statements about the nature of intrinsic goodness.

2. SEARCHING FOR A CRITERION OF INTRINSIC GOODNESS

Here's one way to tackle the question. We can assume that there are several sorts of goodness. First and foremost, of course, there is intrinsic goodness – the "chief good" (in Aristotle's phrase).[5] In addition there are several lesser sorts of goodness. These include goodness as a means, goodness as a sign, goodness as a part, overall goodness, logical goodness (as in "that's a good argument, but the conclusion is false"), monetary goodness, aesthetic goodness, functional goodness and perhaps others. We can then try to determine what's distinctive about intrinsic goodness. We can, that is, seek a "criterion" or "mark" of intrinsic goodness – something that will enable us to distinguish this most important sort of goodness from all the others.

---

[3] Immanuel Kant, *Foundations of the Metaphysics of Morals*, translated, with an introduction, by Lewis White Beck (Indianapolis and New York: Bobbs-Merrill Company, Inc., 1959), p.10.

[4] G. E. Moore, *Principia Ethica*, edited and with an introduction by Thomas Baldwin (Cambridge: Cambridge University Press, 1993): p. 58.

[5] Aristotle, *Nicomachean Ethics*, translated by W. D. Ross, in *The Basic Works of Aristotle*, edited and with an introduction by Richard McKeon (New York: Random House, 1941), p. 941 (Bk. I, Ch. 7. 1097a27).

Several of the great moral philosophers of the past seem to have given some thought to this question, and some of them apparently have provided suggestions of answers. Each of the suggestions embodies a sort of "guiding intuition" about intrinsic goodness. Each suggests one way to distinguish this value concept from others.

I am aware of at least eight main guiding intuitions about intrinsic goodness. (Perhaps it would be better to say that I am aware of some passages in writings of great moral philosophers; if in these passages the philosophers were talking about intrinsic goodness, then these passages seem to contain hints of at least eight distinct guiding intuitions about intrinsic goodness.) I here briefly list the guiding intuitions, give some references, and suggest some sketchy interpretations:

1. The intrinsically good as the unimprovably good: Hints of this intuition can be found in Plato in the *Philebus*, and in Aristotle in the *Nicomachean Ethics* I,7. Aristotle indicates that he is searching for something that is so good that if you have it, your life cannot be improved by the addition of anything else. Happiness (which he takes to be this marvelous thing) is alleged to be "not a thing counted as one good thing among others – if it were so counted it would clearly be made more desirable by the addition of even the least of goods;..." On the contrary, happiness is unimprovably good – it is "...that which when isolated makes life desirable and lacking in nothing." (*NE* I,7)

2. The intrinsically good as the most final good: Again, this idea can be found in Aristotle's *Nicomachean Ethics*. He says, "The chief good is something final... [it is] always desirable in itself and never for the sake of something else." (*NE* I,7)

3. The intrinsically good as the unqualifiedly good: Kant suggests this idea in the *Groundwork*, where he discusses the good will, which according to him shines like a jewel (Section One, first sentence):

> Nothing in the world – indeed nothing even beyond the world – can possibly be conceived which could be called good without qualification except a *good will*...[6] The good will is not good because of what it effects or accomplishes or because of its adequacy to achieve some proposed end; it is good only because of its willing, i.e., it is good of itself.[7]

(The first quoted sentence suggests that intrinsically good things are good without qualification; the second sentence assures us that he is talking about intrinsic goodness.)

4. The intrinsically good as the object of correct intrinsic love: This is Brentano's idea. In *The Origin of our Knowledge of Right and Wrong*,[8] Brentano says: "We call a thing good when the love relating to it is correct. In the broadest sense of the

[6] Kant, op. cit., p. 9.
[7] Kant, op. cit., p. 10.
[8] Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, English edition edited by Roderick M. Chisholm; translated by Roderick M. Chisholm and Elizabeth Schneewind (London: Routledge & Kegan Paul, 1969), p. 18.

term, the good is that which is worthy of love, that which can be loved with the love that is correct."

5. The intrinsically good as that which is good in virtue of its intrinsic nature: This criterion is energetically defended by Moore in "The Conception of Intrinsic Value."[9] Moore says: "We can in fact set up the following definition: To say that a kind of value is intrinsic means merely that the question whether a thing possesses it, and in what degree it possess it, depends solely on the intrinsic nature of the thing in question."

6. The intrinsically good as that which would still be good even if it existed in complete isolation: Moore defends this criterion in many places in *Principia Ethica*. In Chapter VI, for example, he says, "In order to arrive at a correct decision on the question [what things have intrinsic value] it is necessary to consider what things are such that, if they existed by themselves, in absolute isolation, we should yet judge their existence to be good;..."[10]

7. The intrinsically good as the incorruptibly good: Roderick Chisholm in "Objectives and Intrinsic Value" says:

> And what do we mean when we say that a state of affairs is *intrinsically* good, or *intrinsically* bad – as distinguished from being merely *instrumentally* good or *instrumentally* bad? I suggest this: a state of affairs is *intrinsically* good if it is necessarily good – if it is good in every possible world in which it occurs.[11]

Kant seems to have had a similar idea. In his argument about the Good Will, he seems to say that other things (cleverness, courage) are usually good, but can fail to be good under certain circumstances. The Good Will, on the other hand, is good no matter what. Its goodness is "incorruptible."[12]

8. The intrinsically good as that which ought to exist [for its own sake]: Chisholm seems to endorse this idea in "On the Defeat of Good and Evil." After giving a list of things often taken to be intrinsically good, he says, "The things on the good list are the sorts of things that ought to be."[13]

You can almost hear some of these philosophers hyperventilating as they gush with superlatives about intrinsic goodness: "It's a kind of goodness that's unimprovable – if a thing has it, then it's so good that there's nothing you could add

---

[9] "The Conception of Intrinsic Value," in G. E. Moore, *Philosophical Studies* (London: Routledge & Kegan Paul, 1960), p. 260.

[10] Moore, *Principia Ethica*, p. 187.

[11] Roderick M. Chisholm, "Objectives and Intrinsic Value," in *Jenseits von Sein und Nichtsein*, ed. Rudolf Haller (Graz: Akademische Druck- und Verlangsanstalt, 1972), pp. 261-70 [* pp. 171-79 of this volume; the quoted passage appears on p. 172].

[12] Kant, op. cit., pp. 8-9.

[13] Roderick M. Chisholm, "The Defeat of Good and Evil," Presidential Address before the Eastern Division of the American Philosophical Association, *Proceedings and Addresses of the American Philosophical Association* 42 (1968), p. 22.

to it that would make it better." "It's a kind of goodness that's incorruptible – if a thing has it, nothing can detract from its value." "It's a kind of goodness that needs no qualifications – anything that's good in this way is just plain good!"

3. SOME QUESTIONS PROVOKED BY THE GUIDING INTUITIONS

Each of these apparent guiding intuitions about intrinsic goodness has a certain amount of intuitive plausibility, and each also has (to one degree or another) some historical pedigree. Yet each raises lots of questions; and all taken together raise even more questions.

1. First is the question about what the doctrine means. Some of these gushing philosophers wrote long ago and in foreign languages. Some are notoriously obscure writers. Their remarks are sometimes brief and enigmatic. Thus, it is necessary to devote some attention to an initial clarification of meaning.

2. Once the guiding intuitions are clarified, we confront puzzles and paradoxes. Some of these ideas seem rather strange. We have to see if it is possible to refine (or perhaps recast) the ideas, and derive something coherent and useful from them.

3. Even if the intuitions can be recast so as to be somewhat plausible, we face some questions about relations among them. These guiding intuitions might be guiding us toward several different conceptions of goodness. It might turn out for example that Kant focussed on one of these (goodness without qualification) and Aristotle focussed on another (unimprovable goodness). Thus, it might be that the remarks that Kant made about the kind of goodness he had in mind are not equivalent to the remarks that Aristotle made about the kind of goodness he had in mind, and that this is not a mistake. Each of them might have said something true about what he had in mind, but something that might have been false if taken to be about what the other had in mind.

On the other hand, they might both have been talking about the same concept and each of them might have thought that he had found an interesting and distinctive fact about it. If the interesting facts turn out to be equivalent (e.g., a sort of value is unimprovable iff it is unqualified) then that would suggest that Aristotle and Kant were both onto the same sort of value. If the interesting facts turn out not to be equivalent, then several possibilities arise: perhaps one or both of the philosophers was mistaken about a feature of intrinsic goodness; perhaps one or both of them was talking about some other sort of value – a sort of value for which the principle holds true.

4. Suppose the guiding intuitions are not equivalent; suppose they point toward distinct conceptions of intrinsic goodness. Then we face a very deep question: which, if any, of these conceptions are of fundamental importance to moral philosophy?

I would like to answer all of these questions. Obviously, however, it is not possible to discuss them all in one short paper. I am going to focus on just four of these ideas, and just a few of the possible interpretations of each. The four in question are:

1. Unimprovability (Plato and Aristotle): When a thing is intrinsically good, then it's good in such a way as to be "unimprovable" – it can't be made better by the addition of any other sort of goodness. But other sorts of goodness are not in this way unimprovable.

3. Unqualifiedness (Kant): When a thing is intrinsically good, then it is "unqualifiedly good" – you can say that it's good without putting any qualifiers in front of the word "good." But other sorts of goodness are not in this way unqualified. You have to be careful to qualify your remark with phrases such as "so long as it is not being used by a thief or villain."

5. Dependence upon Intrinsic Nature (Moore): When a thing is intrinsically good, its goodness depends upon its intrinsic nature; anything with the same intrinsic nature would have exactly the same amount of goodness. Not so for other sorts of goodness.

7. Incorruptibility (Moore, Chisholm, Kant and others): When a thing is intrinsically good, it is good in such a way that it is good of necessity. It continues to be good in all possible circumstances. Its goodness is incorruptible. Not so for other sorts of goodness.

4. THE APPEAL TO "GOODNESS"

Notice that every one of these alleged marks of intrinsic goodness seems to appeal, on its right-hand side, to some other conception of goodness. Thus, for example, the incorruptibility criterion (7) seems to say that if something is intrinsically good, then it is necessarily "good." The unqualifiedness criterion (3) seems to say that if something is intrinsically good, then the statement that it is "good" needs no qualification. Similar points hold with respect to the other criteria.

Let's start by assuming that when, on the right-hand side of each of these doctrines, we see the word "good," it means "overall good." William Frankena, Chisholm, and others have suggested that the overall value of a thing is the sum of its intrinsic value, its extrinsic value, and any other sort of value it might have.[14] This of course might suggest a sort of circularity in the present context, since we would have to know a thing's intrinsic value in order to perform the addition. Thus, I cannot rely upon it as a definition. Let us assume that a thing is overall good if it's good "all things considered." I hope that the rough idea is familiar enough for present purposes. (It's going to turn out that it does not matter.) Then the doctrines come out meaning this:

[14] William Frankena, *Ethics*, 2nd Edition (Englewood Cliffs: Prentice-Hall, Inc., 1973), p. 83.

1. Unimprovability: When a thing is intrinsically good, then it's *overall good* in such a way as to be "unimprovable" – it can't be made more *overall good* by the addition of any other sort of goodness. But other sorts of goodness are not in this way unimprovable.

3. Unqualifiedness: When a thing is intrinsically good, then it is "unqualifiedly *overall good*." But other sorts of goodness are not in this way unqualified.

5. Dependence upon Intrinsic Nature: When a thing is intrinsically good, its *overall goodness* depends upon its intrinsic nature. Not so for other sorts of goodness.

7. Incorruptibility: When a thing is intrinsically good, it is *overall good* of necessity. It continues to be *overall good* in all possible worlds. Not so for other sorts of goodness.

Every one of these doctrines is obviously false. Consider (1). Surely an intrinsically good thing can be made *better overall* by the addition of other sorts of goodness. Imagine, for example, an episode of pleasure with an intrinsic value of +10 (assuming numerical hedonism for purposes of illustration). The overall value of this episode would be higher if the episode had a lot of good consequences or if it were the sign of good things to come. So it can be improved. That's because the overall value of a thing is affected by its extrinsic value: if we increase the extrinsic value of a thing, then we increase its overall value, even though its intrinsic value remains unaffected.

Consider (5). On the present interpretation, it says that the overall value of an intrinsically good thing depends entirely upon that thing's intrinsic nature. This is obviously wrong. Surely the overall goodness of an episode of pleasure does not depend entirely on the intrinsic nature of that episode. The overall value depends upon a lot of factors including its consequences, what it signifies in its actual context, etc. And just as clearly something else with the same intrinsic nature could differ in overall goodness, since it could differ in extrinsic value.

Consider (7). If (7) were true, then it would be impossible for an intrinsically good thing (e.g., an episode of pleasure) to be overall bad. But all parties agree that no matter how good pleasure is in itself, some episodes of pleasure lead to such bad results that those episodes are not overall good.

Consider (3). If we take "unqualifiedly overall good" to mean "good in such a way that the statement that it is overall good needs no qualifier, but is just fine as it stands," then it is clearly wrong to say that intrinsically good things are unqualifiedly overall good. For an intrinsically good thing might utterly fail to be overall good, and hence it would be wrong to say that it is overall good – whether with a qualifier or not.

These reflections are intended to reveal one central point about the proposed criteria: on the first set of interpretations, every one of them appeals, on its right-hand side, to some concept of goodness. That concept cannot be overall goodness. In order for these criteria to stand a chance of being right, either (a) we must find

some other concept of goodness for the right-hand sides, or else (b) we have to reformulate them entirely, so as to eliminate the need for a distinct concept of goodness on the right-hand side.

5. A DIFFERENT SET OF INTERPRETATIONS

Let's try another set of interpretations of the doctrines. On the right-hand sides of the doctrines I will no longer make any reference to overall goodness. That was a mistake. Consider the following reformulations:

1. Unimprovability: When a thing is intrinsically good, then it has a sort of goodness that is "unimprovable" – in general, and for anything, it can't be made to have more of *that sort of goodness* by the addition of any other sort of goodness. But other sorts of goodness are not in this way unimprovable.

3. Unqualifiedness: When a thing is intrinsically good, then it has a sort of goodness that is "unqualified." In general, and for anything, when you say that it has *that sort of goodness* you don't have to add any qualifier. Your remark will be complete and unambiguous as it stands. But other sorts of goodness are not in this way unqualified.

5. Dependence upon Intrinsic Nature: When a thing is intrinsically good, it has a sort of goodness that depends upon the intrinsic nature of anything that has it. In general, and for anything, if it has *that sort of goodness*, it has it because of its intrinsic nature. Corresponding claims do not hold for other sorts of goodness.

7. Incorruptibility: When a thing is intrinsically good, it has a sort of goodness with respect to which things cannot change. In general, and for anything, if it has *that sort of goodness*, it will continue to have it in all possible worlds. This is not so for other sorts of goodness.

The guiding intuitions are all false on these interpretations, too. In three cases, the doctrine is false because it fails to say something *distinctive* about intrinsic goodness. In one case (unqualifiedness) is goes wrong for another reason.

Consider the first guiding intuition:

1. Unimprovability: When a thing is intrinsically good, then it has a sort of goodness that is "unimprovable" – in general, and for anything, it can't be made to have more of that sort of goodness by the addition of any other sort of goodness. But other sorts of goodness are not in this way self-sufficient.

What's said here about intrinsic goodness is true of many sorts of goodness. Consider signatory goodness for example. Suppose a certain x-ray picture is signatorily good because it indicates good health. Suppose its signatory value is +12. Now add some monetary value, or some instrumental value to the x-ray. You

will find that if it still signifies good health, its signatory value is still +12. Increases in other sorts of value do not serve to increase signatory value.

Similar experiments will show that monetary value, instrumental value, etc. all have the same feature: they are unimprovable in the specified sense. (Overall value is improvable. If you increase the intrinsic value or the extrinsic value of a thing, its overall value will rise.)

3. Unqualifiedness: When a thing is intrinsically good, then it has a sort of goodness that is "unqualified." In general, and for anything, when you say that it has this sort of goodness you don't have to add any sort of qualifier. Your remark will be complete and unambiguous as it stands. But other sorts of goodness are not in this way unqualified.

This guiding intuition seems just plain false. The problem is that "intrinsic" is already a qualifier – it serves to make clear just what sort of goodness is in question. And surely there are plenty of cases in which the failure to include the qualifier would lead to misunderstanding. For example, suppose a certain episode of pleasure causes a dear friend to have a fatal heart attack. The episode of pleasure is extrinsically bad; overall bad; bad as a sign; bad as a part; etc. If you (a hedonist) were to say "that episode of pleasure was good, a 'good thing,' just dandy," you might be misunderstood. It would be better for you to make clear that you mean that it was *intrinsically good* – thus you had better add the qualifier. This shows that the failure to include the qualifier can be just as misleading in the case of intrinsic goodness. Intrinsic goodness cannot be identified as the sort of goodness that needs no qualification.

5. Dependence upon Intrinsic Nature: When a thing is intrinsically good, it has a sort of goodness that depends upon intrinsic natures. In general, and for anything, if it has that sort of goodness, it has it because of its intrinsic nature. But this is not so for other sorts of goodness.

This criterion suffers from a sort of structural problem. On a fairly standard view, the intrinsic nature of a thing is the set of its intrinsic properties. A thing's intrinsic properties are the ones it would have to share with a duplicate. But since duplicates could not differ in intrinsic value, it follows that each thing's intrinsic value is one of its intrinsic properties. As a result, intrinsic values do not "supervene" upon intrinsic natures; they are included within those natures.

Let us bypass this possible difficulty. Let us understand the supervenience criterion broadly, so as to include the possibility that intrinsic goodness is distinctive because it alone of sorts of goodness is either supervenient upon or included within intrinsic natures.

I think it is correct to say that intrinsic goodness supervenes in this way upon intrinsic natures. However, in fact several other sorts of goodness depend in this way upon intrinsic natures, too. Some examples: "logical goodness" as when we say that something is "a good argument"; aesthetic goodness, as when we say that my nose is beautiful; epistemic goodness, as when we say that this is good evidence for that.

Surprisingly, Moore seems to have recognized at least two sorts of intrinsic goodness. In *Principia Ethica*, he clearly assumes that there is just one thing – a "unique property of things" that can be called "[positive] intrinsic value" or "intrinsic goodness."[15] But in "The Conception of Intrinsic Value" he just as clearly indicates that he thinks that there are at least two different sorts of intrinsic goodness. In addition to the former "intrinsic [moral?] goodness" there is beauty, which he takes to be another intrinsic sort of goodness. He casts his "definition" in such a way as to ensure that beauty comes out intrinsic. Although Moore did not say so, logical goodness (validity) comes out intrinsic, as does epistemic goodness (or the goodness of a good reason to believe something).

7. Incorruptibility: When a thing is intrinsically good, it has a sort of goodness that is necessary. In general, and for anything, if it has that sort of goodness then it has it of necessity. That sort of goodness is in this sense incorruptible.

But this feature is not distinctive: several sorts of goodness seem in this way to be necessary (logical, epistemic) and some sorts of goodness that some take to be intrinsic turn out to not be necessary – e.g., beauty.

6. MY INTERPRETATIONS AND EVALUATIONS OF THE INTUITIONS

Here are some rough sketches of what I take to be the most plausible interpretations of the four guiding intuitions:

1. Unimprovability: I think that when Plato and Aristotle spoke of the unimprovability of The Good, they must have been thinking of a possible feature of certain *properties* (e.g., happiness, pleasure, wisdom). They were trying to locate a property they could call "The Good." A person's life would be made better in a distinctive way by the possession of that property. If a person had that property to a sufficiently high degree, then there would be no other property that could improve that person's life. The only thing that could make it better would be getting the first property to a higher degree. When they spoke of a property's being unimprovably good, perhaps that's what they meant.

It seems to me that the quest for an unimprovably good property presupposes a tremendously controversial axiological thesis: that there's *just one* property that's good in the distinctive way. For if there were two, then neither would be unimprovable in this sense. For no matter how much, e.g., wisdom you have, your life could be made better by the addition of some pleasure. Thus it seems to me that the assumption that The Good is unimprovable automatically rules out any interesting form of axiological pluralism.

3. Unqualifiedness: When a thing is intrinsically good, then it is "unqualifiedly good" – you can say that it's good without putting any qualifiers in front of the word "good." But other sorts of goodness are not in this way unqualified.

---

[15] Moore, *Principia Ethica*, p. 68.

Kant might have been trying to state the incorruptibility criterion. If so, the idea would have been that there is no need to add the qualifier because if a thing is intrinsically good, it will continue to be intrinsically good no matter what. So the point is really not that qualifiers are not needed; rather, it is that the object will continue to be intrinsically good no matter what. If this was really Kant's point, he didn't do a very good job of expressing it. On the other hand, if we take him at his word, and understand him to mean that when used to express intrinsic goodness "good" needs no qualifier, his criterion is simply mistaken. Indeed, it would be mistaken no matter what sort of goodness he had in mind. There is no sort of goodness that satisfies it.

5. Dependence upon Intrinsic Nature: When a thing is intrinsically good, it has a kind of goodness that depends upon its intrinsic nature; anything with the same intrinsic nature would have exactly the same amount of that sort of goodness. Not so for other sorts of goodness.

I think it is relatively easy to understand this supervenience criterion. I take it to mean that a sort of goodness is intrinsic iff it either supervenes upon or is included in the intrinsic natures of the things that have it.

This criterion implies that several sorts of goodness are intrinsic – moral, epistemic, logical, aesthetic. It also seems to imply that some things can have their intrinsic values contingently (because some components of intrinsic natures are contingent). My nose formerly had quite a lot of beauty – an intrinsic form of goodness. But now it has slightly less. That's because my nose has changed with respect to some of its intrinsic properties such as shape and size.

If I am right about this, then this guiding intuition fails as a criterion of "that unique property of things" that allegedly stands at the very center of moral philosophy – the one by which ethics can be defined.[16] For if I am right about this, then there are several sorts of intrinsic goodness (aesthetic, epistemic, logical, moral), and some of them have little to do with ethics. A Moorean would then have to say more: what is distinctive about the distinctively *ethical* sort of intrinsic goodness?

7. Necessity; incorruptibility: When a thing is intrinsically good, it has a sort of goodness that things have of necessity. It continues to have just as much of this goodness in all possible worlds and in all possible circumstances. Its goodness is incorruptible. Not so for other sorts of goodness.

This criterion remains problematic. We might take it to mean (a) that intrinsic value is distinctive because if a thing has a certain intrinsic value, it must have that same intrinsic value in all possible circumstances *no matter how different it may be in other respects*. In this case, beauty is not a form of intrinsic goodness, since a thing that is beautiful would fail to be beautiful if were different in its appearance (which it could be). My nose is slightly less beautiful than it formerly was.

On the other hand, we might take it to mean (b) that intrinsic value is distinctive because if a thing has a certain intrinsic value, it must also have that intrinsic value

[16] Moore, *Principia Ethica*, p. 2.

provided that it retains its intrinsic nature. The second interpretation has different implications. Beauty (I think) becomes a sort of intrinsic value. Indeed, several sorts of goodness (logical, epistemic, aesthetic, moral) satisfy the criterion, when stated in this way. Hence, it fails if it is intended as a criterion of the special sort of goodness by reference to which ethics is to be defined.

7. CONCLUSIONS

I have to acknowledge that I have not succeeded entirely in clarifying the four guiding intuitions that I have discussed. Nevertheless, I want to draw some conclusions, making use of rough-and-ready understandings of these things where necessary.

I think Plato and Aristotle's unimprovability criterion is a red herring. They were seeking a mark of The Good; not a mark of intrinsic goodness. If we recast the unimprovability criterion as a criterion of intrinsic goodness as I have suggested, it implies that there can be at most one intrinsically good property. No other criterion has this implication, and it begs too many important questions of axiology. It should be eliminated. (I also think that it does a poor job as a criterion of The Good. My reason for this is the same: it begs too many controversial questions.)

I think the unqualifiedness criterion is simply confused. It implies that there is no such property as intrinsic goodness. It should be eliminated.

I think Moore's supervenience criterion is interesting and fairly plausible, but not as a criterion of that special sort of intrinsic value that interests us as moral philosophers. It allows that there might be several sorts of intrinsic goodness. It also allows that there might be a sort of intrinsic goodness that things have contingently. It rules out the possibility of a sort of intrinsic goodness that things have in virtue of their non-intrinsic properties. As it stands, it is seriously incomplete, since it does not serve to distinguish the relevant sort of intrinsic goodness (moral?) from all the others (aesthetic, epistemic, logical). So, before we appeal to it for crucial projects in moral philosophy (e.g., formulating a principle about right action), we had better complete the project.

I think Kant and Chisholm's incorruptibility criterion allows there to be several sorts of intrinsic goodness. Thus, however plausible we find this as a necessary condition of that central concept of moral philosophy, it is not sufficient. Again, more work has to be done.

Furthermore, this criterion does not permit a sort of intrinsic goodness that things have contingently. It implies that a thing's intrinsic value must be essential to that thing. Some philosophers have thought, however, that the fundamental bearers of intrinsic value are things that happen to be intrinsically good, but might fail under different circumstances to be intrinsically good.

In light of this, it seems to me that we had better be cautious about gushing in this way: "Intrinsic goodness is so wonderful! Things that are intrinsically good have a sort of goodness that is unimprovable! And they have a sort of goodness that is incorruptible!" The problem is that it's not clear that there could be any sort of goodness that's suitably "unimprovable," while there are too many sorts of

goodness that are incorruptible. If we are going to gush, we should at least gush consistently.

In the end, then, it seems to me that of these four guiding intuitions, two are irrelevant and two are incomplete. The incomplete ones do not point in the same direction. Clearly, if we are to answer the complaints of those who are not sympathetic to intrinsic value, we will have to do a much better job of identifying that dazzlingly important object that we think stands at the core of moral philosophy.

APPENDIX

Although I mentioned eight guiding intuitions about intrinsic value, I discussed only four of these in the body of the paper. I here briefly indicate some thoughts about three of the remaining four.

2. Aristotle says that the chief good is "most final," and by this he seems to mean that it is always good as an end, but never good as a means. If taken at face value as a criterion of intrinsic goodness, this remark seems a bit silly. Suppose there are several goods, but happiness is the chief good, as Aristotle seems to say. Surely it still might be the case that happiness is sometimes good as a means. For example, in addition to being outstandingly good in itself, happiness might also be good in virtue of the fact that it lowers stress and thereby leads to good health. It also might be good as a sign, and also good insofar as it serves as a good example to others.

Is it conceivable that Aristotle thought that the chief good would somehow be sullied if it happened to have some good results? Could he have imagined that it would be unfitting for such an august good also to be useful? If so, it appears that Aristotle had a strange conception of the chief good.

6. Moore's isolation test for intrinsic value has been subjected to quite a bit of careful scrutiny.[17] Provided that we take the bearers of intrinsic value to be states of affairs, it turns out to be difficult even to state the test in a coherent way. Aside from whole possible worlds, most states of affairs simply cannot occur without many others occurring alongside. If isolation requires such lonely occurrence, then no such state of affairs can occur in isolation. In this case, the test yields the unfortunate result that everything is intrinsically good.

But as I indicated, the literature on the isolation test is large. A number of intriguing isolationist proposals have been formulated. Many of them have been subjected to critical scrutiny. I cannot review them here.

8. Chisholm suggested in one place that the intrinsically good is that which ought to exist for its own sake. The suggestion is somewhat obscure, since we don't have any agreed conception of what it means to say that something ought to exist for its own sake.

---

[17] See, for example, Eva Bodanszky and Earl Conee, "Isolating Intrinsic Value," *Analysis* 41 (1981), pp. 51-53 [* pp. 11-13 of this volume].

I assume that the intent of the principle is really that the intrinsically good is that which ought to *occur* (rather than *exist*) for its own sake. (I say this because I assume that the candidates for intrinsic goodness are states of affairs. I interpret Chisholm's point as being the claim that good states of affairs are ones that ought to "happen" or occur or be true.)

Consider the state of affairs that consists in my experiencing ten units of pleasure here and now. Call it "F+10." If we assume a simple sort of numerical hedonism, we will want to say that F+10 is intrinsically good. I simply don't know what to say about the question whether F+10 ought to occur for its own sake. Perhaps it ought to occur; but I am inclined to say that if it ought to occur, it ought to occur for *my sake* – because it would make my life go better. I cannot tell whether this runs counter to Chisholm's intent or not.

Suppose I am already experiencing 11 units of pleasure; thus F+11 is occurring. Since the occurrence of F+10 would be slightly worse than what's already happening, it is not entirely clear to me that we should say that F+10 ought to occur for its own sake.

PART II

DOUBTS ABOUT THE CONCEPT OF INTRINSIC VALUE

CHAPTER 7

M. BEARDSLEY

INTRINSIC VALUE

1.

Many philosophers apparently still accept the proposition that there is such a thing as intrinsic value, i.e., that some part of the value of some things (objects, events, or states of affairs) is intrinsic value. John Dewey's attack seems not to have dislodged this proposition, for today it is seldom questioned. I propose to press the attack again, in terms that owe a great deal to Dewey, as I understand him.

The predicates (1) "…has intrinsic value," (2) "…is intrinsically valuable," and (3) "…is intrinsically desirable," will be used interchangeably – not for the sake of elegant variation, but because each permits idiomatic constructions that bring out different features of what I take to be the same concept. "Desirable" associates with "to desire," which will be convenient to have available at a later stage of the argument. (No doubt "valuable" and "to value" are similarly related, but the latter is not free from the suggestion of reflective appraisal.) On the other hard, the noun "value" is useful because we can speak of "a value" and "a kind of value." "Value" lends itself more readily than "desirability" to such adjectival qualifications as "cognitive value" and "moral value."

Two phrases are the most often used in defining "intrinsic value": "for its own sake" and "in itself." Their meanings are close, but not identical, and the second seems more satisfactory than the first.

We might say that something is intrinsically valuable, in some degree, if it is valuable for its own sake, and that if it has value for the sake of something else, then its intrinsic value, if any, is that which would remain if that other-regarding value were subtracted. One inconvenience of this definition can be brought out as follows: A sheet of postage stamps has been misprinted – the central figure, say, is inverted. The stamps derive part of their value from their rarity. Is one of these stamps valuable, in part, for its own sake? Well, its value is not for the sake of anything else – if we speak of its philatelic value, not its market value. But is this value then intrinsic? It seems strange to say this when it can be taken away, without altering the

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 61-75.
© 2005 *Springer. Printed in the Netherlands.*

stamp at all, simply by having the Post Office Department print a few hundred million more copies. Since its rarity is a relational property, there is a sense in which the rare stamp is valuable not for the sake of anything else, either. It might be replied that, even if the issue becomes plentiful, the philatelic value of each individual stamp is not destroyed, but only reduced; rarity cannot transform an object with no value into one with value; it can only increase certain sorts of value in things that already have some degree of it. Still, that part of the stamp's philatelic value that is supplied by rarity seems to be neither intrinsic nor extrinsic, if these are defined as "for its own sake" and "for the sake of something else" respectively.

The second definition of "intrinsic value" is that proposed by G. E. Moore in his paper on "The Conception of Intrinsic Value." Suppose we can distinguish between the internal and external properties of a thing, that is between (1) its qualities and inner relations, and (2) its relations to other things. Then the value that depends upon a thing's internal properties alone is its intrinsic value; the value that depends (wholly or partly) upon a thing's external properties is its extrinsic value. The intrinsically good thing is "good in itself." Moore states the definition this way:

> To say that a kind of value is "intrinsic" means merely that the question whether a thing possesses it, and in what degree it possesses it, depends solely on the intrinsic nature of the thing in question.[1]

It is this definition of intrinsic value that leads to Moore's thought-experiment in *Principia Ethica.* To decide the question "What things have intrinsic value?", he says,

> it is necessary to consider what things are such that, if they existed *by themselves*, in absolute isolation, we should yet judge their existence to be good.[2]

If the intrinsic value of a thing is independent of its relationship to anything else, it cannot be destroyed by the removal of everything else. Moore holds that only by applying this test can we sort out intrinsic from extrinsic value with clarity and confidence.

Some puzzles in Moore's definition of "intrinsic nature" have been further discussed by him and by others.[3] One difficulty is to explain how intrinsic goodness can be a property dependent solely on the intrinsic properties of things, without itself being an intrinsic property – this explanation is needed to fix the status of goodness as a "nonnatural property." Another is to decide whether dispositional properties are to be called internal or external, and to dispose of a certain element of arbitrariness or conventionality in the classification of some properties as dispositional. In the case of the most self-contained and self-sufficient of the valuable objects we are acquainted with – that is, works of art – the internal-external distinction has an immediate appeal. Thus we tell people to listen to the music "itself" and pay no

[1] *Philosophical Studies*, London 1922, p. 260.
[2] *Principia Ethica*, Cambridge University 1903, p. 187; cf. *Ethics*, London 1947, pp. 42, 101.
[3] See C. D. Broad, "G. E. Moore's Latest Published Views on Ethics," *Mind* LXX (1961), pp. 435-57.

attention to anything outside it, such as objects it might suggest or the biography of the composer. And the so-called “Formalist” has been known to assert that the (aesthetic) value of a painting depends on internal properties alone (lines, shapes, and colors) and owes nothing to its representational relationship to the world outside it. But even if these notions are regarded as acceptable, serious questions can be raised about the sharpness and decisiveness of the internal-external distinction.

And it is no doubt for this reason, as well as for others, that most contemporary value-theorists have concluded that if anything is intrinsically valuable it is not an external object, but an experience or psychological state. At best, the work of art could only be said to have “inherent value” (in C. I. Lewis’s terminology), if exposure to it can result in an aesthetically enjoyable experience. The distinction between internal and external properties seems clearer when applied to experiences. Of course, intentionality – a reference to other states of mind or to the external world – must be taken as internal to the experience itself; but neither an ostensible memory of a past pleasure nor the expectation of a future one is causally dependent upon the occurrence of those pleasures, and so, theoretically at least, we can conceive any short stretch of conscious life apart from its antecedents and consequents. And apparently we can ask whether it has intrinsic value.

I take it that to say that something is valuable is to say that it deserves to be valued; and to say that something is desirable is to say that it is worthy of being desired. Now when we add “intrinsically” to, say, “desirable,” how does it fit into the definiens? Does “X is intrinsically desirable” mean

(1) “X is intrinsically worthy of being desired” (that is, by definition, “X is worthy-of-being-desired on account of its internal properties alone”)?

Or does it mean

(2) “X is worthy of being intrinsically desired” (that is, by definition, “X is worthy of being-desired-on-account-of-its-internal-properties-alone”)?

I have puzzled over the relationship between these two expressions, and find that I can only understand the former in terms of the latter. For if X’s desirability depends on its internal properties alone, then these must be the properties that ought to be, or deserve to be, desired; and what ought to be the case is that X is desired on account of these internal properties.

Thus in order to attach a sense to “intrinsically desirable” we must first attach a sense to “intrinsically desired.” And there is, I think, no trouble about this. For there is an evident psychological distinction between desiring something on account of its internal properties alone, and desiring it on account of its relationships to other things. The distinction is not easy to apply, because in most of our desires, care for the thing itself and concern for what will come of it are thoroughly mixed. But we can pretty well fix the extremes: at the one end, the candy that the child in the

grocery store wants, and screams for, and evidently would count the world well lost for; at the other end, the pieces of string and bits of cellophane that we instantly discard, once we have secured the goods they serve to wrap. And we can use Moore's test with some success to make the psychological distinction, by asking a person, for example, to think what he would choose to do if he had only a short time on earth, with limitless resources and no obligations to others. The child no doubt would gorge himself on candy, but nobody would pore over chewing-gun wrappers.

So I would like to allow, and exploit, the term "to desire intrinsically." But of course it does not follow automatically that because we can attach "intrinsically" to "desire," we can therefore attach it to "desirable." There are obviously other adverbs that go with "desired," but not with "desirable" (for example, "eagerly," "strongly," and "widely"). An argument must be made out to show that things can be intrinsically desirable as well as intrinsically desired.

2.

The question, then, is this: What good reasons are there, if any, for believing that there is such a thing as intrinsic value? Since this proposition is seldom considered to be in need of elaborate proof, formal arguments are difficult to collect. I can only discover three such arguments: (1) an argument from definition, (2) a dialectical demonstration, and (3) an attempt at empirical confirmation.

(1) *The Argument from Definition.* In the view of some thinkers, the existence of intrinsic value can be simply shown in this way: Some extrinsic value is instrumental value, which is defined as follows:

> "X has instrumental value" means "X is conducive to something that has intrinsic value." (Call this Definition A)

I have selected the loose term "conducive to," in order to avoid some distinctions that we do not need at present. Thus if Y is an end to which X is a means, or a whole of which X is a necessary part, then X is conducive to such intrinsic value as Y possesses. Obviously, if we accept Definition A, we are as committed to the existence of intrinsic value as we are to the existence of instrumental value – however long the chain of conduciveness may be.

But must we accept this definition? In order for Y to confer its value on X, when X is conducive to Y, it is certain that Y must have some value to confer, but whether that value is intrinsic or instrumental does not matter as far as X is concerned. So the following alternative definition should be acceptable:

> "X has instrumental value" means "X is conducive to something that has value." (Call this Definition B)

The Arguer from Definition rejects Definition B. It is all right if the key term in the definiens, "value," can be defined by itself, without reference to intrinsic value. But the word "value," he might contend, is necessarily an ellipsis; it cannot stand by

itself. Up to this point I have been speaking as though value is a genus with two species, so that value can be defined first and then divided by Moore's test. And it is true that in describing that test, Moore speaks as though we could first know that an object has a certain total value, before going on to discover how much of that total is intrinsic, and how much extrinsic. But this, according to the Arguer from Definition, is all misleading. The terms "extrinsic value" and "intrinsic value", despite the noun they share, do not name coordinate species of a genus, but designate two very different concepts, one of which is derivative from the other.[4] And Moore's own way of speaking agrees with this interpretation, on occasion. For example, he speaks[5] as though the phrase "good as a means" (i.e., having instrumental value) is synonymous with the phrase "a means to good" (i.e., conducing to intrinsic value). Being good as a means is not a way of being good – the instrumentally valuable thing is not a valuable instrument, strictly speaking, but an instrument of value. And similarly,

> To have value merely as a part is equivalent to having no value at all, but merely being a part of that which has it.[6]

This line of thought would issue in the rejection of my Definition B. For it seems to show that instrumental value can only be defined in terms of intrinsic value; so that the existence of the former automatically entails the existence of the latter.

Nevertheless, it seems to me that the word "value" does have a meaning by itself, and does mark out a genus. We can sensibly ascribe value as such to things, and this is even more clearly true of "desirable" and "worth having."[7] The terms "good as a means" and "good as an end in itself," if they may be taken as synonymous with "instrumental good" and "intrinsic good," suggest that the distinction is between two *grounds* of goodness, not two *senses.* This is the position assumed by Moore in his *Ethics* when he asserts that

> saying a thing is intrinsically good...means it would be a good thing that the thing in question should exist, even if it existed *quite alone*, without any further accompaniments or effects whatever.[8]

For here "intrinsically good" is defined by means of "a good thing." If this is a correct procedure, Definition B is acceptable, and the Argument from Definition fails.

(2) *The Dialectical Demonstration* is closely connected with the Argument from Definition, but deserves exhibition on its own. I think of it as logically parallel to the First Cause argument for the existence of God. "Instrumentally valuable" is a

---

[4] "Intrinsic" and "extrinsic," in the convenient terminology of Austin Duncan-Jones, would then be "sense-discriminating," rather than "concept-modifying," adjectives (see "Intrinsic Value: Some Comments on the Work of G. E. Moore," *Philosophy* XXXIII (1958), pp. 240-73, esp. pp. 261-62).

[5] *Principia Ethica*, p. 24.

[6] *Ibid.*, p. 35.

[7] See Duncan-]ones, pp. 257-58.

[8] *Ethics*, p. 42.

relational concept – X borrows its value from Y, or Y confers its value upon X. If the value Y confers is itself instrumental, so that it is merely passed along from Z, then where does Z get its value? In the last analysis, something must (according to this argument) possess its value in itself, or nothing can get any value.[9] So the existence of any instrumental value proves the existence of some intrinsic value, just as the occurrence of any event is said to prove the existence of a First Cause.

To align the Demonstration of intrinsic value with the venerable First Cause argument may lend it prestige, but may also suggest its faults. As Kant showed, the First Cause argument projects a certain kind of ideal explanation that cannot be completed if the causal series has no beginning term. That is, if to explain an event, X, requires not only that we assign its cause, W, but that we assign an *explained* cause, then the ideal explanation of X would involve the explanation of all its causal antecedents; and if these have no first term, no such explanation can be given. Similarly, the Dialectical Demonstration of intrinsic value projects a certain kind of ideal justification that cannot be completed if the series of means and ends has no last term. That is, if to justify ascribing a value to X requires not only that we show it is a means to Y, but also that we justify ascribing a value to Y, and if there is no stopping point, no such ideal justification can be given. But ordinary justification, like ordinary causal explanation, involves no such infinite regress.

The Dialectical Demonstration cannot be a pure formal demonstration, for, unless the Argument from Definition is valid, it cannot be proved strictly self-contradictory to assert the existence of instrumental value but deny the existence of intrinsic value. The Demonstration must rather be thought of as applying to our *knowledge* of value. Premise 1: We know, or have good reason to believe, that some things are instrumentally valuable. Premise 2: We could not know this unless we knew some things to be intrinsically valuable. Conclusion: We know some things to be intrinsically valuable. Now, Premise 1 seems to me clearly true. But the conclusion seems to me quite clearly false. The paradoxical feature of our value-knowledge is just that we have a good deal of sound knowledge about instrumental values, but are in considerable doubt about intrinsic values. Philosophers have disputed, and still dispute, about whether pleasure is an intrinsic value, and, if so, whether it is the only intrinsic value; and it is significant that as ordinary people we have not had to wait upon the settlement of these issues before discovering a great many valuable things nearer to hand. We must have some way of knowing that in many concrete situations it is better for a person to be healthy than sick, without knowing whether that is because it is intrinsically best for him to maximize his net positive hedonic quality, or realize his potentialities, or cultivate a good will, or whatever.

The apparent hopelessness of resolving problems about instrumental value without knowing antecedently what, if anything, intrinsic value may be, can be

[9] Cf. Hume's argument that "something must be desirable on its own account," *Inquiry Concerning the Principles of Morals*, N.Y.: Liberal Arts Press, 1957, p. 111.

cleared away, I believe, if we go back to the concrete contexts of value-problems.[10] When we are in the position of having to decide what is valuable, or more valuable, we are in Dewey's "problematic situation," and such a situation is one in which certain ends are in grave doubt and others are (on that occasion) taken as temporarily fixed. If the value of everything in the situation were in question at once, nothing could be decided at all, and indeed no problem could even be conceived; only in terms of certain tentatively-held values, can we decide, or even ask, whether other things are valuable or not. There must be a basis on which it is reasonable for us to pick out salient elements in the situation, and assign them probable values, without transforming the task of assignment into another problematic situation, thus endlessly postponing a decision. A state of affairs such as good health, for example, has retained its eligibility through earlier problems and experiences; its value has survived them, in the sense that up to this point health does not seem to have interfered with our pursuits, but a lack of it from time to time has not only contributed to the rise of difficult problems, but has limited our capacity to resolve them. We need not suppose that health is an intrinsic good, or that it is always good, or even that we are necessarily right in taking it to be good on the present occasion – but we do have rational justification for supposing that it has positive value that ought to be taken into consideration now.

This is the merest sketch of a way of looking at the problem of evaluation. The gist of it is that reasonable decisions about instrumental values do not presuppose, or wait upon, previous reasonable decisions about intrinsic values (even if such decisions were possible, which I shall argue – in the next section – they are not). So there is no infinite regress in a purely instrumentalist theory of value.

(3) *The Empirical Confirmation.* Some writers on the theory of value have not relied on either of the a priori forms of argument that I have just considered, but have held that the existence of intrinsic value is attested by direct experience. One of the most carefully considered theories is that of C. I. Lewis. He distinguishes three kinds of value-predication, of which the basic kind is the "expressive statement," having a form like "This is good," where "this" refers to immediate presentations in experience.[11] Expressive statements report "value-apprehensions" or "direct findings of value-quality in what is presented,"[12] and are therefore incorrigible.[13]

> ... It will hardly be denied that there is what may be called "apparent value" or "felt goodness," as there is seen redness or heard shrillness. And while the intent to formulate just this apparent value-quality of what is given, without implication of anything further, encounters linguistic difficulties, surely it will

[10] Here I want to acknowledge my debt to Sidney Hook's essay on "The Desirable and Emotive in Dewey's Ethics," in his collection, *John Dewey: Philosopher of Science and Freedom*, N. Y. 1950.
[11] *Theory of Knowledge and Valuation*, LaSalle, Ill., 1946, p. 374.
[12] *Ibid.*, p. 365.
[13] *Ibid.*, p. 375.

> not be denied that there are such immediate experiences of good and bad to be formulated.[14]

It is this immediate value-quality that Lewis calls intrinsic value, and that imports intrinsic value into those experiences that possess it. To it all forms of value-judgment are ultimately anchored. And because it is only in experiences that we are directly acquainted with such qualities, only experiences can, strictly speaking, have intrinsic value. When an object enables us to "realize" intrinsic value in experience, by being directly presented to us, the object has "inherent value," which is a form of extrinsic value.[15] "The goodness of a good *object* is a potentiality for the realization of goodness in experience."[16]

> We have suggested – and intend to abide by – distinction between intrinsic and extrinsic values by reference to the question, "Is that which is valued, valued *for its own sake* or for the sake of something else?"[17]

What is most surprising in Lewis's characterization of this value-quality, is the way he employs alternately, and indifferently, the two expressions that other philosophers have been at such pains to keep distinct: "valuable in itself" and "valued in itself." There are passages in which Lewis seems to suggest that being valuable and being valued may not be the same thing, since one of them may be evidence of the other.

> Such appellations as these – "liked" or "disliked," "wanted" or "unwanted," "good" or "bad," as addressed to the directly presented – are better indices of the immediately valuable or disvaluable than others.[18]

Here being liked is an *index* of being intrinsically valuable, hence not identical to it (though the force of this distinction is certainly much blunted by putting "good" and "bad" in with "liked" and "disliked"). And later he says, in the antecedent of a concessive conditional, that "liking and disliking are decisive of immediate value,"[19] again as though there might be an evidential relationship between being immediately liked and being intrinsically valuable. But in Lewis's prevailing usage, "valuable" is simply reduced to "liked," or "likeable," and "*per se* desirable" only means "*per se* desired." The normative element in "value" is completely lost sight of.

It is intrinsic desiredness that Lewis substitutes for intrinsic desirability. Thus he speaks of "the likeability and dislikeability of things; their directly gratifying quality or the opposite,"[20] and says that "value-disvalue is that mode or aspect of the given or the contemplated to which desire and aversion are addressed."[21] These remarks

---

[14] *Ibid.*, pp. 374-75.
[15] *Ibid.*, pp. 386-87, 391.
[16] *Ibid.*, p. 389.
[17] *Ibid.*, p. 385.
[18] *Ibid.*, p. 404.
[19] *Ibid.*, p. 410.
[20] *Ibid.*, p. 418.
[21] *Ibid.*, p. 403.

suggest that when he speaks of "desirability," he means only "likeability," and when he speaks of "finding value-qualities" he means only finding something that excites desire or aversion. But in that case, Lewis has not really shown that the existence of intrinsic desirability (in the usual sense) is confirmed in immediate experience.[22]

Charles A. Baylis[23] explicates "intrinsic good" as what is "worthy of existence entirely apart from any extrinsic value it may have" (an interesting example of reversing the order of definitions proposed in the Argument from Definition above). And he explicates "worthy of existence" by saying

> that anything which has this characteristic ought, *ceteris paribus*, to exist rather than not, that it would be better for it to exist, and that anyone who can bring it into being ought to do so unless there is something preferable he can do instead.[24]

Like W. D. Ross, in short, he makes "intrinsic value" entail a conditional ought-sentence, a prima facie obligation. He then goes on to say that "the best initial evidence ... we could have" for ascribing intrinsic value to something

> is that we find ourselves *prizing* things of that kind, *i.e.*, liking, approving, desiring, preferring, and commending them, for their own qualities (rather than because of their relations to other valuable things) in circumstances where to the best of our searching knowledge we are making no mistake in our cognition of them. Such evidence gives us an initial probability that what we thus prize is intrinsically good.[25]

Since Baylis agrees with Lewis that only experiences can be intrinsically good, his position contrasts interestingly with Lewis's on the question of our knowledge of intrinsic value. For Lewis, the desiring of a mental state, while it is occurring (i.e., the enjoyment of it), is conclusive evidence of its desirability. For Baylis, it is only "initial evidence," the probative force of which is then to be increased

> by making repeated examinations of things of the same kind under circumstances which vary just enough to guard against the kinds of cognitive error which might occur.[26]

It seems to me that Baylis gives two completely unrelated accounts of intrinsic value, and only makes them seem the same by passing so swiftly from one to the other. Up to the middle of his paper, all characterizations of intrinsic value are, in terms of words like "worthy," "better," and "ought" – and it clearly retains its normative character. But when he comes to his thesis that particular well-conducted

---

[22] Cf. Duncan-Jones: "I do not think that worth-havingness is more or less the same as pleasantness or enjoyableness" (*op. cit*., p. 266), because "worth having," involves an element of "deliberation." And cf. John Dewey: "To pass from immediacy of enjoyment to something called 'intrinsic value' is a leap for which there is no ground" (*Theory of Valuation*, International Encyclopedia of Unified Science, Vol II, No. 4, Chicago 1939, p. 41).

[23] "Grading, Values, and Choice", *Mind* LXVII (1958): 485-501.

[24] *Ibid*., p. 494.

[25] Baylis, *op. cit*., pp. 494-95.

[26] *Ibid*., p. 495.

prizings constitute *evidence* for intrinsic value, a quite different characterization takes over.

> The attribution of intrinsic goodness to an experience is like the attribution of an ideally defined, but not actually observable, physical property, *e.g.*, weighing precisely one pound, to an object.[27]

Weighing exactly one pound is explicated in terms of being equal in weight to a standard pound; and the latter turns out to be a complicated conditional about what would happen under ideal weighing conditions (admitting that the ideal conditions cannot all be specified). Thus "X weighs exactly one pound" means (approximately) "X would be found to be one pound in weight by an Ideal Weigher, under Ideal Conditions, with an Ideal Scale, etc., etc." And similarly, "X has intrinsic value" turns out to be equivalent to a complex conditional to the effect that X would be prized by an Ideal Observer, defined by all the necessary accoutrements and qualifications of such an individual.[28]

I do not propose to take up the Ideal Observer theory here, or related proposals for defining intrinsic value, such as have been very carefully worked out by Roderick Firth and Richard Brandt.[29] Yet I hesitate to set aside so important a position without more careful consideration. Essentially, my view of it can be put in the form of a dilemma. When "X has intrinsic value" is defined in terms of the attitude, or desires, or satisfactions, of an Ideal Observer, the specified characteristics of the Ideal Observer either do or do not include any normative concepts. If the Ideal Observer is defined wholly in nonnormative terms, as one who is omniscient, impartial, etc., then (1) statements about the Ideal Observer can be confirmed by reports of my own prizings, under controlled conditions, but (2) it will remain an open question, in Moore's sense, whether the Ideal Observer actually desires what is desirable; and from the statement that something is desired by the Ideal Observer, nothing follows about what deserves to exist, or to be done. On the other hand, if normative terms are smuggled into the definition of the Ideal Observer – so that he is defined in effect as one who knows intrinsic value when he sees it – then (1) it does follow that I ought to desire, as far as I can, what the Ideal Observer desires, but (2) statements about the desires of the Ideal Observer can no longer be empirical hypotheses, and cannot be confirmed, or given initial probability, by reports of any actual desirings or prizings; from the statement that a certain musical experience is found prizeworthy by an Ideal Observer, no predictions can be derived about my prizings or disprizings.

---

[27] *Ibid*., p. 497.
[28] *Ibid*., p. 499.
[29] See Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research* XII (March 1952): 317-45; Firth and Brandt, "The Definition of an 'Ideal Observer' Theory in Ethics," *ibid*. XV (March 1955): 407-23; Richard Henson, "On Being Ideal," *Philosophical Review* LXV (1956): 389-400; and Richard Brandt, *Ethical Theory*, Englewood Cliffs, 1959, ch. 10.

I conclude that the existence of intrinsically valuable things, including mental states, cannot be made certain by direct experience or probable by inductive inference from direct experience.

It is interesting, by the way, that Paul Taylor, in his recent book, rejects the Argument from Definition, and holds that "a world where all values were extrinsic, the value of one thing depending on the value of another whose value in turn depends on the value of something else, *ad infinitum*," is not logically impossible, nor even unimaginable.[30] In his view, it is an empirical fact about our world that all extrinsic values ultimately do go back to intrinsic ones. But it seems to me that Taylor, for all his care, mixes the desirable and the desired as Lewis and Baylis do. Thus, in describing the "world in which instrumental and contributive values did *not* depend on intrinsic value," he says that in such a world

> No one would do anything for its own sake, simply because he found personal enjoyment in it. It would be a world of "practical people" who knew how to get things done but had no reason for getting one thing done rather than another.[31]

But this is a caricature of a world in which all values are extrinsic. For, in the first place, the question whether people "do anything for its own sake" is not at all the question whether there is intrinsic value; the world without intrinsic value (that is, our world) is a world in which people may often do something for its own sake, and may experience many enjoyments, but they cannot *justify* doing something for its own sake by simple appeal to enjoyment. And in the second place, a world in which all values are instrumental would be precisely a world in which every correct value-judgment could be supported by a reason, and so there would always be a "reason for getting one thing done rather than another."

3.

I now turn from the arguments for the thesis that there exists at least one thing with intrinsic value, to my argument against this thesis. It is that the concept of intrinsic value is inapplicable – that even if something has intrinsic value, we could not know it, and therefore that it can play no role in ethical or aesthetic reasoning.

Richard Brandt has remarked that "X is desirable" means the same as "Desiring X is justified."[32] What "desirable" adds to "desired" is this claim to justifiability. But the only way this claim can be made good is by considering X in the wider context of other things, in relation to a segment of a life or of many lives. Thus the term "intrinsic desirability" pulls in two directions: the noun tells us to look farther afield, the adjective tells us to pay no attention to anything but X itself.

---

[30] *Normative Discourse*, N.Y. 1961, p. 26.
[31] *Ibid.*, p. 32.
[32] *Op. cit.*, p. 302.

This implicit contradiction in the concept of intrinsic value can, in theory, be removed on one epistemological assumption. If we could detect the presence of intrinsic value by immediate intuition – in the way Moore claimed to know that the satisfactions of personal relationship and of artistic beauty are the highest intrinsic values available to human beings[33] – then we would not require reasons in order to know it. If we ask why the experience of a good painting has intrinsic value, the answer will be "It contains (or is characterized by) aesthetic enjoyment."[34] And if we then ask why aesthetic enjoyment is intrinsically valuable, we have reached an end. For no further reasons can be given, so long as we are confined, in answering, to the experience itself. We can only say that it is self-evident to some cognitive faculty that is equipped to grasp the self-evidence of such truths as this. I cannot honestly appeal to such intuitive apprehensions of intrinsic value, when I do not possess them myself. But I can fairly question whether others are mistaken in thinking that they possess them. The familiar arguments against ethical intuitionism will not be reviewed in detail here. Since intuition is a last resort as a way of resolving the problems of value theory, it is fatal that ethical intuitions conflict; that criteria cannot be given for distinguishing correct from incorrect ones; and that the alleged analogies with other types of *a priori* knowledge (mathematical, for example) break down. It can never be a necessary truth that any particular instance of aesthetic enjoyment is desirable, since later experience may reveal that it was not in fact desirable at all; therefore it can never be a necessary truth that any particular instance of aesthetic enjoyment is *intrinsically* desirable; therefore no *general* or *universal* proposition about the connection of aesthetic enjoyment and intrinsic value can be necessary; and therefore, there is no necessary truth about the intrinsic value of aesthetic enjoyment to be grasped by intuition.

Brandt gives an excellent example for discussion:

> Consider a child who is swinging, in a rapturous state of enjoyment. We shall probably think that being in this state of mind (and perhaps body) is worthwhile for itself alone. To be in a state of rapturous enjoyment of the experience of swinging is for one's state of mind to have an intrinsic property, on account of which the child's experience is desirable. So we shall say that the child's experience is of intrinsic worth.[35]

Now, how could one justify the assertions that this experience is (a) intrinsically desirable and (b) desirable? If intrinsic desirability is in question, we can only say, "It is an experience containing the rapturous enjoyment of swinging." But it would be odd if we were content to accept this as a sufficient reason for saying that the experience is intrinsically desirable, when we would not accept it as a sufficient reason for saying that the experience is desirable. For to support the latter statement, we would need at least to say that in the first place the swinging doesn't do the child

---

[33] *Principia Ethica*, p. 188.

[34] Supposing there is such a special sort of enjoyment; see my paper on "The Discrimination of Aesthetic Enjoyment," *British Journal of Aesthetics*, III (1963), pp. 291-300.

[35] *Op. cit.*, p. 303.

any harm (it is an innocent pleasure), and that there is nothing else more important for him to be doing at this time. Otherwise we should not say that this particular experience of swinging is desirable at all. But as soon as we bring in these considerations making the desirability of this experience depend in part upon the lack of unwanted consequences and upon the comparison of this action with other possible actions, then the desirability we have defended is no longer intrinsic, but extrinsic.

When we call a thing "desirable," it seems to me, we claim to place it in a larger perspective; we suggest that alternatives have been considered, and the desirable thing selected with a view to its implications and connections. And this is the trouble with Moore's test, and all such artificial desert-island ethical models. Let us suppose the child has only a few minutes to live, because the world is about to come to an end, and he is rapturously swinging; and suppose anything else you like, to close off other alternatives. He wants to swing, he is ecstatic about it. Of course we are not going to call him in to do his homework, or worry about his catching cold. Suppose we narrow his opportunities to two, swinging or doing some unpleasant alternative, like washing the dishes. We are asked to choose which he shall do, and we say, "Let him swing!" Are we then conceding that the enjoyment of swinging has intrinsic value? I do not think so.

It is essential to distinguish between two propositions about this one-minute-to-live version of Moore's test:

(1) There is no reason for the child not to swing on this occasion, if he wants to and enjoys it.

(2) There is reason to say that the child's swinging enjoyment on this occasion is intrinsically desirable.

The situation we have artificially set up is designed to make the first proposition true. But this is a very different proposition from the second, which is what the example is supposed to prove.

What can be shown by the swinging-example is that we can imagine cases in which there could be nothing to make a particular experience undesirable – neither intrinsic nor extrinsic reasons. But once we deliberately set aside all extrinsic factors, there is nothing to argue about, either for or against. No reason can be given why the child should not swing if he wishes, and since he does not need a reason for doing what he already wants to do, reasons do not come into the picture at all. In normal everyday life, with its continuities and connections, the fact that he desires to swing would raise the question of its desirability – but it would not settle that question, of course. It would also assign the burden of proof. When the child expresses, or evinces, the desire to swing, the next question is whether any reasons can be given against the satisfaction of this desire (of course, if such reasons are given against it, it will then be in order to see whether reasons can be given *for* it). What makes the artificial situation artificial is just that it rules out any possible reasons against swinging, and so prevents any real problem from arising. I accept

the spirit (though not every word) of a remark by William James, which is quoted with approval by William Frankena in his recent book on ethics:

> Take any demand, however slight, which any creature, however weak, may make. Ought it not, for its own sole sake, to be satisfied? If not, prove why not.[36]

If not, why not? That is exactly what I mean by the burden of proof.

And I think that this concept of the burden of proof is very important in many problems. We say that in a criminal trial the burden of proof lies upon the prosecution. An a prioristic philosopher might express the situation this way: "In criminal trials, there is an antecedent probability that the accused is innocent." Then the question would arise in the philosophy of law: how can that antecedent probability of innocence be established? And hopeless problems would appear. Is it a pragmatic postulate? Is it self-evident? Is it intuited by a special sense of justice? Is it deducible from theological premises about the nature of man? But all these questions would be beside the point. And I think the same thing holds for intrinsic value. The child wants to swing. It is a mistake to say that this fact gives initial probability to the proposition that the pleasure of swinging has intrinsic value. And it is a mistake to say that the pleasure he gets in swinging *justifies* the act of swinging. Say rather that his wanting to swing raises the question whether it would be justified for him to swing, but as long as no reason can be given *against* it, the swinging does not *require* any justification.

The value-judgment, "This act of swinging is desirable," is best thought of as an answer to a question, a solution of a problem. But so long as the child wants to swing, and in the nature of the case there can be no objection, there is no problem to be solved. To say that "beauty is its own excuse for being" is to say that beauty *needs* no excuse for being.

The artificial dead-end situation dreamed up according to Moore's instructions is essentially misleading. But some real-life situations may seem to approximate it. There are occasions when we want to say things like: "You'll find Mammoth Cave well worth seeing in itself," or "Go ahead and eat the ice cream cone; it's tasty," or "Forget business. Relax. Just have a good time." Borrowing a Deweyan term, let us call these "consummatory judgments" – they point out that something can be enjoyed for its own sake, they locate sources and occasions of intrinsic enjoyment in food, art, etc. They do not seem to be judgments of extrinsic value: they are hedged about in such a way as apparently to recommend a course of action without considering consequences at all. Are they, then, really judgments of intrinsic value?

I think this inference would be mistaken. Consummatory judgments are better interpreted not as value-judgments but as statements to the effect that no judgment is required, because there is no conflict of values, no occasion for deliberation and choice. "Go ahead and swing" surely has to be based on a preliminary survey of the situation: the rope looks safe, the child's stomach is not easily upset; there is no

---

[36] *Essays in Pragmatism*, N. Y. 1948, p. 73; see William Frankena, *Ethics*, Englewood Cliffs, 1963, p. 38.

homework in the offing; there are no visiting children who should have the first turn on the swing, etc. Once we know these facts we can adopt a vacation point of view, and, when the child asks politely, or simply starts running toward the swing, we say, "Go ahead." This is not a judgment of intrinsic value at all, but a kind of *nihil obstat*. If some contrary reason loomed, we would have to stop and think, and decide. Consummatory judgments are then judgments of a metalinguistic order. "No reason why you shouldn't if you want to," is the basic formula.

In closing, I should like to pay my respects to John Dewey once more. I am always frustrated in reading Dewey, trying to separate the enormously good points from the confusing ones. Much of Dewey's famous attack on intrinsic value is really concerned with something else, namely ends-in-themselves (as opposed to ends-in-view). What he exposes over and over again is the danger of fixing on goals without reasonable regard to their means and consequences, and he is convinced that the belief in intrinsic value fosters this fixation, with its attendant train of ills: fanaticism, utopianism, opportunism, and the rest. But of course it does not logically follow that if there are intrinsically valuable things then there are necessarily ends-in-themselves. The world might be full of things that have intrinsic value, but whose intrinsic value is always outweighed by harmful effects. This would be the opposite of F. H. Bradley's description of Leibniz's universe, when he said it was the best of all possible worlds but everything in it was a necessary evil. The means might be delightful, but the ends always awry. It is only in Moore's artificial isolation that a thing with intrinsic value would necessarily be an end, and a great part of Dewey's concern should have been allayed by assurance that from the fact that a thing would be an end in isolation it does not follow that it would be an end in the context of human life. Nevertheless, I believe Dewey was also right in rejecting intrinsic value, and that this part of his theory has important ethical and social consequences, too. To connect the desirable and the desired, to connect values with human needs and wishes, is indeed the task of a naturalistic theory of value. But to make this connection prematurely, through an identification of intrinsic value with immediate enjoyments, encourages a dangerously one-sided approach to human problems. For then desires, which provide the data and conditions of value-problems, and set some of the limits within which solutions are to be found, are taken as something more final than they really are, rather than as states of affairs that may themselves have to be transformed. And reasoning about values may become too exclusively a matter of balancing interests, and finding ways to satisfy existing and conflicting wants, rather than an inquiry into their conditions in the natural and social environment. When we recognize the essential relatedness of all value, this helps to remind us that in the deepest and most poignant conflicts in ourselves or in our society, it is often desires themselves that must be transformed if lives are to be freed and fulfilled.

CHAPTER 8

C. M. KORSGAARD

TWO DISTINCTIONS IN GOODNESS

1.

In this paper I describe two distinctions in goodness which are often conflated, and try to show the importance of keeping them separate. The two distinctions in question are: the distinction between intrinsic and extrinsic goodness, and the distinction between ends or final goods, and means or instrumental goods.

It will help to begin by delineating the kind of value and the kind of judgment of value with which I am primarily concerned here. I take it that there are three primary categories of value with which the moral philosopher is concerned: namely, the rightness or justice of actions, policies, and institutions; the goodness of objects, purposes, lives, etc.; and the moral worth or moral goodness of characters, dispositions, or actions. My concern here is not with what constitutes moral worth or moral goodness but with the second category – with goodness as a feature of ordinary ends and purposes, states of affairs, objects, activities, and other things – that is, with the kind of goodness that marks a thing out as worthy of choice.

Within this category, we can distinguish, admittedly with some artificiality, three kinds of judgments of goodness that we make. We judge something to be good of its kind when we judge it to have the virtues appropriate to that kind. We may also judge something to be a good kind of thing, as when we say of friendship or books or health that they are good. And we also sometimes judge particular things to be good absolutely, meaning that here and now the world is a better place because of this thing. I am mostly concerned with this third sort of judgment in this paper, though part of what is in question is its relation to the other two.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 77-96.
© 2005 *Springer. Printed in the Netherlands.*

2.

It is rather standard fare in philosophy to distinguish two kinds of this value of goodness, often called "intrinsic" and "instrumental."[1] Objects, activities, or whatever, have an instrumental value if they are valued for the sake of something else – tools, money, and chores would be standard examples. A common explanation of the supposedly contrasting kind, intrinsic goodness, is to say that a thing is intrinsically good if it is valued for its own sake, that being the obvious alternative to a thing's being valued for the sake of something else. This is not, however, what the words "intrinsic value" mean. To say that something is intrinsically good is not by definition to say that it is valued for its own sake: it is to say that it has its goodness in itself. It refers, one might say, to the location or source of the goodness rather than the way we value the thing. The contrast between instrumental and intrinsic value is therefore misleading, a false contrast. The natural contrast to intrinsic goodness – the value a thing has "in itself" – is *extrinsic* goodness – the value a thing gets from some other source. The natural contrast to a thing that is valued instrumentally or as a means is a thing that is valued for its own sake or as an end. There are, therefore, two distinctions in goodness. One is the distinction between things valued for their own sakes and things valued for the sake of something else – between ends and means, or final and instrumental goods. The other is the distinction between things which have their value in themselves and things which derive their value from some other source: intrinsically good things versus extrinsically good things. Intrinsic and instrumental good should not be treated as correlatives, because they belong to two different distinctions.

If intrinsic is taken to be the opposite of instrumental, then it is under the influence of a theory: a theory according to which the two distinctions in goodness are the same, or amount to the same thing. According to such a theory, final goods or things valuable as ends will be the same as intrinsic goods, and instrumental goods or things valuable as means will be the same as extrinsic goods. It is worth considering what such a theory might be like.

The first part of the equivalence – that ends and intrinsic goods are the same – might be held in two very different ways: (I) The claim might be that anything we value for its own sake is thereby "intrinsically good"; that is, that this is all that can be meant by "intrinsically good." This amounts to a *reduction* of the intrinsic/extrinsic distinction to the end/means distinction; the significance of the former distinction drops out. This option, which in effect replaces the intrinsic/extrinsic distinction with the end/means distinction, is sometimes taken to render the conception of good "subjective," both in the sense of "relative to the

[1] Intrinsic is often directly opposed to instrumental: equally commonly, "extrinsic" is opposed to intrinsic but then "consequential" or "instrumental" is offered as a definition or explanation of that term. Or, in some of the literature, "intrinsic" is taken to be a particular theory about how ends are valued, and accepted or dismissed as such. All of these usages more or less imply the equivalence of the two distinctions; none leaves room for the Kantian theory described in this paper.

person" and of "varying among individuals." The thought goes this way: good things (on this account) are just those valued for their own sakes, but different people value different things for their own sakes. (2) The second way one might equate ends and intrinsically good things is by claiming that those things which have intrinsic value are or ought to be treated as ends. In this case we have a significant, and rather metaphysical, claim about ethics and moral psychology: namely, that choice is or ought to be a response to an attribute that we perceive in things – the attribute of intrinsic goodness. Equating the two distinctions in goodness thus leads naturally to the idea that there are two alternative theories about final goods – either that "good" is subjective or that good things are the possessors of some particular attribute. Objectivity, in other words, is thought to amount to the possession of an attribute. I think that many people do have a tendency to think that these alternatives are exhaustive, and one thing I want to show is that if the two distinctions in goodness are kept separate, this need not be so.

The other side of the theory that equates the two distinctions is the equation of extrinsic with instrumental goods, or means. The consequences of this equation are serious. Since intrinsically good things (at least when "intrinsic" retains its significance) are thought to have their value *in* themselves, they are thought to have their goodness in any and all circumstances – to carry it with them, so to speak. If you find that a certain kind of thing is not good in any and all circumstances, that it is good in some cases and not in others, its goodness is extrinsic – it is derived from or dependent upon the circumstances. If extrinsic value and instrumental value are equated, you are then forced to say of all such things that they are means or instruments. This way of thinking is part of what is behind the tendency to conclude that the final good must be pleasure or some sort of experience. The argument proceeds as follows: take an activity that we would naturally say is valuable for its own sake – say, looking at a beautiful sunset. Now the question is raised: would you think this activity was a good one even if the person engaged in it found it tedious or painful? If you say "no" then you have admitted that the goodness of this activity is not intrinsic; that it depends, in some way, on the pleasantness of it. But if all extrinsic value is instrumental value, then the only option is that the activity is a *means* to pleasure. Now if the two distinctions are not equated there is room for some other sorts of accounts of extrinsic value, and one may not be forced to this conclusion.

Because of these consequences this side of the equation has been more widely attacked than the other. It has been argued that instrumentality is not the only sort of extrinsic value, on the grounds that there are other sorts of contributions things can make to intrinsically good ends. So, for instance, it is common to identify a "part" of an intrinsically valuable "whole" as having "contributive" value. Another sort of value, suggested by C. I. Lewis, is called "inherent" value.[2] This is supposed to be the value that characterizes the object of an intrinsically good experience. A painting, for example, might have inherent value. The identification of these

[2] C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle: Open Court, 1946).

different kinds of extrinsic value serves as a reminder that things can bear other relations to good ends besides being their causes or tools for their production. Contributive value and inherent value, however, both share with instrumental value the fact of deriving their goodness from the contribution they make to the existence of a supposedly intrinsically good end.

Separating the two distinctions in goodness, however, opens up another possibility: that of something which is extrinsically good yet valued as an end. An example of this would be something that was good as an end because of the interest that someone took in it, or the desire that someone had for it, for its own sake. This is the case that I am going to be discussing in the rest of this paper. In particular, I am going to compare the very opposite treatments of this issue that appear in, on the one hand, the work of Moore and Ross at the beginning of our century and, on the other hand, Kant. These philosophers all separated the two distinctions, but they applied them to this case in opposite ways. Moore and Ross came to the conclusion that the goodness of ends is intrinsic and must be independent of the interest that people take in them or the desires that people have for them. You might value something as an end because of its intrinsic goodness or in response to its intrinsic goodness, but a thing's possession of intrinsic goodness is quite independent of whether anyone cares about it or not. Kant's theory, on the other hand, both allows for and depends upon the idea of extrinsically valuable ends whose value comes from the interest that people take in them.

The fact that philosophers nowadays often oppose intrinsic to instrumental value and equate intrinsic value with the value of ends may just be taken to be sloppiness, of course. But it may also mean that these philosophers are working with some theory of the sort I have described – a theory of the equivalence of the two distinctions. As the Kantian option shows, such a theory is a substantive philosophical position and restricts the possibilities open to us in serious ways. It should not, in any case, be taken for granted.

3.

In the early years of this century there was much discussion of the question whether or not a good thing has its value as a result of something like the interest taken in it or the desire someone has for it. Influential philosophers such as G. E. Moore, W. D. Ross, R. B. Perry, and others discussed this question at length. Probably the interest in the issue was aroused by a common utilitarian argument that pleasure is the only thing that is good in itself because it is the only thing that we can desire for its own sake. It quickly became, and still is, a commonplace in discussions of utilitarianism to argue that pleasure is not, after all, the only thing that we desire for its own sake. But that leaves open the further question whether the things we desire for their own sakes, whatever they might be, are therefore good in themselves, or intrinsically good. Moore, and following him, Ross, argued vigorously that this could not be so. Goodness, they said, had nothing to do with mental attitudes taken towards things at all – even though it turned out that, as a matter of fact, goodness is a property of

mental attitudes or a property of states of affairs that always include mental states or attitudes.

The idea of intrinsic value is central to Moore's theory. He believed that right actions are those that maximize intrinsic goods. Emphatically opposed to hedonism, he took the class of intrinsic goods to consist of such things as the appreciation of beauty, friendship, and love. In his attempt to account for the goodness of these things, he came back to the question of the nature of intrinsic value over and over again.[3]

In his paper "The Conception of Intrinsic Value," Moore argues that people who object to the idea that goodness is subjective are really worried about something quite different: the idea that goodness is nonintrinsic. This is shown, according to Moore, by the fact that there are theories which would render goodness objective to which the same people would still be opposed, and for the same reason. Moore gives as his example the theory that "better" means "better fitted to survive":[4] people who object to subjectivity, he says, would also object to this, although it renders "good" objective. So the problem with a subjectivist theory of the good is not merely the lack of objectivity, but something else. According to Moore, it is that it excludes the possibility that things are intrinsically valuable. Moore defines intrinsic value as follows:

> To say that a kind of value is "intrinsic" means merely that the question whether a thing possesses it, and in what degree it possesses it depends solely on the intrinsic nature of the thing in question.[5]

Moore's definition of the intrinsic nature of a thing is rather complicated: he says that two things have a different intrinsic nature if they are not exactly alike; that the difference need not be a difference in qualities, since it may be in the degree of a quality or in the quality of a constituent; and that two numerically different things have the same intrinsic nature if they are exactly alike.[6] In general, the intrinsic nature of a thing seems to consist of its nonrelational properties, for Moore insists that a thing would have the same intrinsic nature if transferred to another world or placed in a different set-up of causal laws.[7] This is what Moore supposes we *want* from the conception of intrinsic goodness, as his analysis of the trouble with the evolutionary account of goodness shows. He says that the difficulty is that the types better fitted to survive under our laws of nature would not be the same as the types

---

[3] Moore's views on intrinsic value are mostly presupposed in *Principia Ethica* (Cambridge, Cambridge 1903); but they are addressed explicitly in *Ethics* (Oxford, 1912); "The Conception of Intrinsic Value" written for *Philosophical Studies* (Routledge and Kegan Paul Ltd., 1922); and a symposium reply entitled "Is Goodness a Quality?" published in the *Aristotelian Society Supplement* in 1932 and reprinted in *Philosophical Papers* (1959). In the last, Moore tends to give way to a view that his earlier accounts avoid – namely, that only experiences can be intrinsically good. For that reason and because of its polemical nature I have not used it in this paper.

[4] "The Conception of Intrinsic Value" p. 256.

[5] *Ibid*., p. 260.

[6] *Ibid*., pp. 260-65.

[7] *Ibid*., p. 256.

better fitted to survive under other circumstances and with different laws of nature. "Good" therefore would not be dependent only on a thing's intrinsic nature but would be a property that is relative to the circumstances, even though in this case it would be objective. But the problem with subjectivism is the same: it makes "good" relative to the circumstances.

Intrinsic goodness is not an element in the thing's intrinsic nature, for to say that would be to commit the naturalistic fallacy. The elements in its intrinsic nature are natural properties and cannot be identified with the good. But it is dependent only on the thing's intrinsic nature and is just as constant: so long as the thing remains what it is, it has the same value: and the value is the same, of course, for everyone and so also objective. Since it is no part of a thing's intrinsic nature whether anybody likes it or not, intrinsic value is quite independent of people's desires and interests. To put it another way: the attribute of "being desired by somebody" is relational, and as such it obviously varies with the circumstances in which the thing is found.

In *Ethics*, Moore's definition is a little different. We judge a state of things to have intrinsic value when we judge that it would be a good thing for that state of things to exist, even if nothing else were to exist besides. Here again, the emphasis is on the thing's goodness being nonrelational in a certain way. This view of intrinsic goodness is behind Moore's method of determining which things have intrinsic goodness in *Principia Ethica*: the "method of isolation." In order to arrive at a correct decision on the question which things have intrinsic value, Moore says that we must consider whether a thing is such that, if it existed by itself, in absolute isolation, we should judge its existence to be good.[8] In *Ethics*, Moore says:

> We *can* consider with regard to any particular state of things whether it would be worth while that it should exist, even if there were absolutely nothing else in the Universe besides... [W]e *can* consider whether the existence of such a Universe would have been better than nothing, or whether it would have been just as good that nothing at all should ever have existed.[9]

These definitions, along with the method of isolation they suggest, seem to Moore to exclude easily any connection between intrinsic value and what people desire for its own sake, for, he tells us, it is obviously possible to desire something for its own sake, or believe that someone else does, and yet not regard the thing as the sort of thing that would be good if it existed in isolation. Indeed you might regard it as a bad thing, worse than nothing, for it to exist quite alone. Moore concludes:

> And if this is so, then it shows conclusively that to judge that a thing is intrinsically good is not the same thing as to judge that some man is pleased with it or desires it for its own sake.[10]

Moore, it should be noted, does not usually use the terminology of "relational" vs. "nonrelational" attributes in his discussions of intrinsic value, but these are the

---

[8] *Principia Ethica*, p. 187.
[9] *Ethics*, p. 68.
[10] *Ibid*., p. 69.

terms in which Ross and Perry, following Moore, take up the discussion. Ross, who is on Moore's side,[11] says that there are two kinds of theories of value. One kind treats value as an attribute, and the other treats it as a relation, usually to a state of mind such as interest or desire. If it is a relation, Ross complains, then nothing can be intrinsically good, since intrinsically good *means* "good even if nothing else exists." But, he says,

> ... in that case value would seem always to be borrowed, and never owned; value would shine by a reflected glory having no original source.[12]

Ross, like Moore, finds it virtually self-evident that "intrinsically valuable" is not the same as "desired as an end." He insists that:

> It is surely clear that when we call something good we are thinking of it as possessing in itself a certain attribute and are not thinking of it as necessarily having an interest taken in it.[13]

The terms in which this discussion proceeded suggested that the question was whether final goods, whatever we ought to pursue, are intrinsically good and objective, the possessors of a property; or good because they are desired and therefore subjective, or at any rate "relational" and therefore unfixed. These are terms that those who followed Moore and Ross inherited.

4.

Kant, I am going to claim, was aware of and made use of the two distinctions in goodness, with results that were quite different from those arrived at by Moore and Ross.[14] In order to see this, we must begin by looking at Kant's own distinction between unconditioned and conditioned value. The *Foundations of the Metaphysics of Morals* opens with the famous claim:

---

[11] In *The Right and the Good* (Oxford, 1930) Chapter IV, Ross argues explicitly in favor of Moore and against Ralph Barton Perry, who, in his *General Theory of Value* (Harvard, 1926) argues that value is relative to interest. The Kantian view defended in this paper is classified by Perry as one in which value is "the object of a qualified interest" and opposed by him in favor of the view that value is "the object of any interest."

[12] *The Right and the Good*, p. 75.

[13] *Ibid.*, p. 81.

[14] I am not the first to set up Kant's view in opposition to Moore's. The same is done by H. J. Paton in "The Alleged Independence of Goodness" written for the *Library of Living Philosophers* volume on Moore (Northwestern: Volume 4, 1942). Paton, however, is not concerned with the two distinctions, and he focuses on the goodness of actions, which he claims is relative to the circumstances in which they are performed. Moore's rather impatient response is to incorporate the choice into the action and consideration of the circumstances into the choice: thus under different circumstances you have different actions. Moore's reply may be fair in the case discussed, but it is an instance of a general strategy which I discuss in the paper: when someone brings forward an example of a good thing whose goodness seems relative to the circumstances, Moore and Ross incorporate the circumstances into the thing to maintain the nonrelational character of the goodness.

> Nothing in the world – indeed nothing even beyond the world – can possibly be conceived which could be called good without qualification except a *good will.* (G 9/392-93)[15]

As Kant presents the argument that follows, it becomes clear that what he means is that the good will is the only unconditionally good thing and "the supreme condition to which the private purposes of men must for the most part defer" (G 12/396). He says:

> This will must indeed not be the sole and complete good but the highest good and the condition of all others, even of the desire for happiness. (G 12/396)

Happiness, by contrast to the good will, is referred to as a "conditional purpose" (G 12/396).

The fact that happiness is identified as a conditional purpose shows that the unconditioned/conditioned distinction is not the same as the end/means distinction, since happiness is certainly desired as an end. For Kant, the end/means distinction can be said to be a distinction in the *way* we value things. By contrast, the unconditioned/conditioned distinction is a distinction not in the way we value things but in the circumstances (conditions) in which they are objectively good. A thing is unconditionally good if it is good under any and all conditions, if it is good no matter what the context. In order to be unconditionally good, a thing must obviously carry its own value with it – have its goodness in itself (be an end in itself). Kant's notion of unconditional value therefore corresponds to the notion of intrinsic goodness as nonrelational that I have been discussing. The early passages of the *Foundations* emphasize the independence of the value of the good will from all surrounding circumstances as well as from its results. It is good in the world or even beyond it (G 9/393); it is not good because of what it effects or accomplishes; it sparkles like a jewel in its own right, as something that has its full worth in itself. Later in the *Foundations,* Kant uses the phrase "*inneren Wert*," inner worth, to describe the special dignity of a morally good rational being as compared to the "*relativen Wert*," relative worth, of anything else (G 53/435). But whereas Moore assigned intrinsic goodness to a range of things – to aesthetic appreciation, to friendship, and in general to the things that he thought we ought to pursue as ends – Kant assigns it only to the one thing, the good will.

If unconditional value is intrinsic value, conditional value is extrinsic value. Now a thing is conditionally valuable if it is good only when certain conditions are met; if it is good sometimes and not others. Thus, to elaborate on Kant's own examples, "the coolness of a villain makes him not only far more dangerous but also more directly abominable in our eyes than he would have seemed without it" (G 10/344), while coolness in a fireman or a surgeon is usually an excellent thing. Power, riches,

---

[15] References to Kant's works are given in the text as shown. "G" stands for the *Foundations of the Metaphysics of Morals*; the first page number is that of the translation by Lewis White Beck (Library of Liberal Arts, 1959); the second is the Prussian Academy edition page number. Other titles are given in full. The translations used are: Lewis White Beck, *Critique of Practical Reason* (Library of Liberal Arts, 1956); and Mary J. Gregor, *The Doctrine of Virtue* (Harper Torchbooks, 1964).

and health are good or not depending upon what use is made of them. To say that a thing is conditionally valuable is to say that it is good when and only when the conditions of its goodness are met. We can say that a thing is good objectively (this is my terminology) either if it is unconditionally good or if it is a thing of conditional value and the conditions of its goodness are met. Here it is important to notice that "good objectively" is a judgment applying to real particulars: this woman's knowledge, this man's happiness, and so on. To say of a thing that it is good objectively is not to say that it is the type of thing that is usually good (a good kind of thing like knowledge or happiness) but that it contributes to the actual goodness of the world: here and now the world is a better place for *this.* We would not say that about the coolness of the villain or the happiness of the evil person: hence coolness and happiness are objectively good only when certain further conditions are met. Further, we might, under unusual conditions, attribute objective goodness to something that under more usual conditions is nearly always bad, as when a kind of occurrence normally unfortunate coincidentally contributes to someone's happiness.[16] When Kant says that the only thing good without qualification is a good will, he means that the good will is the one thing or kind of thing for which the whole world is always a better place, no matter "what it effects or accomplishes" (G 10/394).

The two distinctions interact in the following ways. When a thing is valued as a means or instrumentally (or is the sort of thing valued as a means) it will always be a conditionally or extrinsically valuable thing, and the goodness of the end to which it is a means will be a condition of its goodness. Instruments therefore can only be conditionally valuable. If the conditions of their goodness are met, however, they can be good objectively. The more important point is about things valued as ends. These are also conditionally or extrinsically good. In particular, happiness, under which Kant thinks all our other private purposes are subsumed, is only conditionally good, for:

> It need hardly be mentioned that the sight of a being adorned with no feature of a pure and good will, yet enjoying uninterrupted prosperity, can never give pleasure to a rational impartial observer. Thus the good will seems to constitute the indispensable condition even of worthiness to be happy. (G 9/393)

But although happiness is conditionally valuable, it is, when the condition is met, objectively good.

In order to see this, it will help to keep in mind Kant's other uses of the unconditioned/conditioned distinction. If anything is conditioned in any way, reason seeks its condition, continually seeking the conditions of each condition until it reaches something unconditional. It is this characteristic activity of reason that generates the antinomies of theoretical speculative reason described in the *Critique of Pure Reason.* The usual example is causal explanation – if we explain a thing in

[16] I am indebted to the Editors of the *Philosophical Review* for this point.

terms of its cause, we then go on to explain the cause itself in terms of its cause, and this process continues. Reason does not want to rest until it reaches something that needs no explanation (although this turns out not to be available): say, something that is a first cause or its own cause. A causal explanation truly satisfying to reason would go all the way back to this evident first cause, thus *fully* explaining why the thing to be explained must be so. These are familiar sorts of moves in philosophy, so there is no need to belabor the point. To apply it here, it is only necessary to point out that just as to explain a thing fully we would have to find its unconditioned first cause, so to *justify* a thing fully (where justify is "show that it is objectively good") we would have to show that all the conditions of its goodness were met, regressing on the conditions until we came to what is unconditioned. Since the good will is the only unconditionally good thing, this means that it must be the source and condition of all the goodness in the world; goodness, as it were, flows into the world from the good will, and there would be none without it. If a person has a good will, then that person's happiness (to the extent of his or her virtue) is good. This is why the highest good, the whole object of practical reason, is virtue and happiness in proportion to virtue: together these comprise all ends that are objectively good – the unconditional good and the private ends that are rendered good by its presence (*Critique of Practical Reason*, 114-115/110) So also the Kingdom of Ends, defined as "a whole of rational beings as ends in themselves as well as of the particular ends which each may set for himself" (G 51/433), is a kingdom in which the objectively good is fully realized.

On the Kantian conception of goodness, then, an end is objectively good either if it is unconditionally (intrinsically) good *or* if it is conditionally good and the relevant conditions, whatever they are, are met. This conception of the good is used in his argument for one of the formulas of the categorical imperative, the Formula of Humanity as an End in itself.[17] It is this argument that establishes the role of the good will in conferring value upon the ends of the person who has it.

The argument shows how Kant's idea of justification works. It can be read as a kind of regress upon the conditions, starting from an important assumption. The assumption is that when a rational being makes a choice or undertakes an action, he or she supposes the object to be good, and its pursuit to be justified. At least, if there is a categorical imperative there must be objectively good ends, for then there are necessary actions and so necessary ends (G 45-46/427-28; *Doctrine of Virtue* 43-44/384-85). In order for there to be any objectively good ends, however, there must be something that is unconditionally good and so can serve as a sufficient condition of their goodness. Kant considers what this might be: it cannot be an object of inclination, for those have only a conditional worth, "for if the inclinations and the needs founded on them did not exist, their object would be without worth" (G 46/428). It cannot be the inclinations themselves because a rational being would rather be free from them. Nor can it be external things, which serve only as means.

[17] A much fuller treatment of the ideas of this section is in my paper "Kant's Formula of Humanity," *Kant-Studien* 77 (1986), pp. 183-202.

So, Kant asserts, the unconditionally valuable thing must be "humanity" or "rational nature," which he defines as the capacity to set an end (G 56/437; DV 51/392). Kant explains that regarding your existence as a rational being as an end in itself is a "subjective principle of human action." By this I understand him to mean that we must regard ourselves as capable of conferring value upon the objects of our choice, the ends that we set, because we must regard our ends as good. But since "every other rational being thinks of his existence by the same rational ground which holds also for myself" (G 47/429), we must regard others as capable of conferring value by reason of their rational choices and so also as ends in themselves. Treating another as an end in itself thus involves making that person's ends as far as possible your own (G 49/430). The ends that are chosen by any rational being, possessed of the humanity or rational nature that is fully realized in a good will, take on the status of objective goods. They are not intrinsically valuable, but they are objectively valuable in the sense that every rational being has a reason to promote or realize them. For this reason it is our duty to promote the happiness of others – the ends that they choose – and, in general, to make the highest good our end.

It is worth emphasizing that the relation of intrinsic to extrinsic value in this case – the case of extrinsically valuable ends – is entirely different from that in the cases of extrinsic value mentioned earlier. Instrumental value, contributive value, and Lewis's inherent value were all forms of extrinsic value that derived from the production of a supposedly intrinsically good end. The extrinsic value of an objectively good end – of something that forms part of the happiness of a good person – comes not from some further thing that that end promotes but from its status as the object of a rational and fully justified choice. Value in this case does not travel from an end to a means but from a fully rational choice to its object. Value is, as I have put it, "conferred" by choice. This formulation may seem paradoxical. A natural objection will be that the goodness of the chosen object is precisely what makes the choice rational, so that the choice cannot itself be what makes the object good. I will have more to say about this objection in the next section. The point I want to emphasize here is that the Kantian approach frees us from assessing the rationality of a choice by means of the apparently ontological task of assessing the thing chosen: we do not need to identify especially rational ends. Instead, it is the reasoning that goes into the choice itself – the procedures of full justification – that determines the rationality of the choice and so certifies the goodness of the object. Thus the goodness of rationally chosen ends is a matter of the demands of practical reason rather than a matter of ontology.[18] It is notable that on Kant's theory the goodness of means is handled the same way: it is not because of the ontological property of being productive of an intrinsically good end that means are good but rather because of the law of practical reason that "whoever wills the end, so far as

[18] Insofar as Moore's point in identifying the naturalistic fallacy is to deny the identity of goodness with any particular natural property and so to insist on the autonomy of ethical discourse, Kant could agree. But whereas Moore concludes that goodness must therefore be a nonnatural property, Kant understands it to be a practical, rather than a theoretical, characterization.

reason has decisive influence on his action, wills also the indispensably necessary means to it that lie in his power" (G 34/417). Similarly, the argument for the objective goodness of the object of a rational choice is not an ontological one; rather, it is based on Kant's theory of rational action. If we regard our actions as rational, we must regard our ends as good; if so, we accord to ourselves a power of conferring goodness on the objects of our choice, and we must accord the same power – and so the same intrinsic worth – to others.

It will be helpful to pause for a moment to match up Kant's view and the Kantian terms to what has gone before. On Kant's view there is only one thing that has what he calls unconditional value and what Moore calls intrinsic value, and that is the power of rational choice (when the choices are made in a fully rational way, which is what characterizes the good will). The value of everything else whatever is extrinsic or conditional. Yet when a thing is conditionally valuable and the relevant conditions are met, the thing has objective value. Things that are valued for their own sakes or as ends have this status. Their value is conditional but can be objective, given the real circumstances of the case. Thus, although Kant, like Moore, firmly separates intrinsic value from a thing's being desired for its own sake, he has resources for saying that a thing is objectively good as an end *because* it is desired for its own sake. And most things that are good will in fact be good in this way: they will be good because they are part of the happiness of a deserving human being. On Kant's theory, the goodness of most things is, in the way described by Ross, relational – relative to the desires and interests of people. But since it must also be appropriately related to one thing that has intrinsic value, it is not merely "subjective." Value does, in Ross's extravagant terms, "shine with a reflected glory," and it is "borrowed rather than owned" by most of the things that have it. But it does have an original source that brings it into the world – the value-conferring power of the good will.

5.

In this section I want to focus on some advantages of the Kantian way of describing values. In the next section I will show how some of these advantages are shared by Moore. In the last section I will discuss what I take to be the most important advantage that Kant's theory of goodness has over Moore's.

Kant's treatment of the two distinctions and the relations between them allows us to describe certain kinds of everyday matters of value in a way that is more flexible and that I think is more natural than is available to us if the two distinctions are conflated or equated. This is especially so for certain cases of what we might call "mixed" values. I have in mind a variety of different mixtures. Take some examples: a luxurious instrument; a malicious pleasure; an unenjoyed exercise of one's higher faculties; or an undisplayed art object. Now the idea that a thing can have value under a condition, when combined with the reminder that instrumentality or usefulness is not the only possible condition (that is, some extrinsically good things are valued as ends), will help us to describe such cases.

Consider, for instance, a common symbol of aspiration – a mink coat. Is it valuable as a means or as an end? One hardly wants to say that it is valuable only as a means, to keep the cold out. The people who want mink coats are not willing to exchange them for plastic parkas, if those are better protection against the elements. A mink coat can be valued the way we value things for their own sakes: a person might put it on a list of the things he always wanted, or aspires to have some day, right alongside adventure, travel, or peace of mind. Yet it is also odd to say it is valued simply for its own sake. A coat is essentially instrumental: were it not for the ways in which human beings respond to cold, we would not care about them or ever think about them. To say that the coat is intrinsically or unconditionally valuable is absurd: its value is dependent upon an enormously complicated set of conditions, physiological, economic, and symbolic. Certainly, it does not pass Moore's isolation test, so far as I can see. A universe consisting of a mink coat or of someone's having one, without the associations that can only be provided by the particular relations and causal connections under which we live, is not really imaginable, much less valuable. What would a coat *be*? It seems hard even to apply the isolation test here, for one is tempted to say that its instrumentality *is* one of the elements in the "intrinsic nature" of a coat, even though it can hardly be said to be a property the coat would have under any set of laws of nature. If its instrumentality is not one of its intrinsic properties, then one is regarding the coat as something else – an animal skin sewed into a peculiar shape, perhaps. But then it seems as if one must strip away the practically relevant properties of the coat in order to ask about its intrinsic value – and that cannot be right. It is equally absurd to say of such a thing that it is a mere instrument, just because its value is conditioned. The Kantian distinctions allow us to say that the coat is valued in part for its own sake, although only under certain conditions. It even allows us to say of certain kinds of things, such as luxurious instruments, that they are valued for their own sakes under the condition of their usefulness. Mink coats and handsome china and gorgeously enameled frying pans are all things that human beings might choose partly for their own sakes under the condition of their instrumentality: that is, given the role such things play in our lives.

Another possible advantage is that the independent use of the two distinctions will provide us with a way of talking about the relation of pleasure, enjoyment, and appreciation to other kinds of value that does not turn these mental states into ends to which everything else is a means. Activities of various kinds might be thought to be good under the condition that we enjoy them and not good at all for those who, for one reason or another, cannot enjoy them, without forcing the conclusion that it is only for the sake of the enjoyment that they are valued. Certain difficulties concerning the "higher pleasures" described by Mill or those activities that Aristotle says are "pleasant in their own nature" although not necessarily "to a particular person" might be dealt with in this way. But this is a suggestion I cannot pursue here.

Consider also the example of an extraordinarily beautiful painting unsuspectedly locked up, perhaps permanently, in a closet. Now a beautiful painting, I am

supposing, is valued for its own sake. If the two distinctions are equated, we must say it has intrinsic value. Yet it is locked in a closet, utterly unseen, and no one is the better for its existence. Consider Moore's isolation test: is a universe with such a painting locked up somewhere intuitively better than one without it? Is a universe consisting of such a painting better than a universe consisting of something quite plain, with no viewers in either? These are curious puzzles: and Moore's isolation test seems to force us to ask the metaphysical-sounding question whether the painting has this property, intrinsic value, or not. Yet we *know* what the practically relevant property of the painting is: it is its beauty. Now on the Kantian type of account we can say that the painting is valuable for its own sake, yet so long as it remains locked up and unseen, it is no good at all. The condition of its goodness – the condition of the goodness of its beauty – is not met. That condition is that the painting be viewed. Yet although its value is not intrinsic, the painting may be objectively good for its own sake. If it were viewed, and the viewer were enraptured, or satisfied, or instructed by its loveliness, then the painting would be an objectively good thing: for the world would be, really, a better place for it: it would be a substantive contribution to the actual sum of goodness of the world. Notice, too, that this does not in the least mean that we have to say that the painting is only valued as a means to the experiences of appreciation. Those experiences are not an end to which the painting is a means, but the condition under which its value as an end is realized.

I am not suggesting that the Kantian treatment solves all the difficulties in our thinking about these things, but only that it does not drive us immediately to the conclusion that all of these things, valued only under conditions and only in a network of relations, must be mere instruments or contributors to some further thing – pleasure or some "mental" state, which supposedly has the real value. The conflation of the two distinctions does tend to have this effect. In particular, when conflation leads us to the conclusion that a thing can only be valued as an end when it is intrinsically valuable, or valuable independently of all conditions and relations, we find ourselves led inevitably to the curious conclusion to which modern moral philosophers are indeed frequently led – that everything good as an end must be something mental, some kind of experience. I have already mentioned one line of argument that leads to this conclusion: some sort of experience, such as pleasure, seems to be a condition of goodness of so many good things. Another line of thought that leads this way is this: no matter how much the philosopher wants to insist that the value of a good thing must be intrinsic and so nonrelational, the sense remains that the goodness of a good thing must have something to do with its goodness *for us.* It cannot merely be a property, metaphysical and simple, which we perceive in things and respond to in an extraordinary way. So the fact that goodness must lie in some relation to human beings, evidently at odds with the theory that goodness must be entirely nonrelational, is dealt with by making goodness a property of something belonging directly to the human being – our experiences or states of mind. By making goodness lie in the experiences themselves, the philosopher rids us of the worry: but what if no one is around to care about this good thing? What good is it

then? Kant's way of looking at it, on the other hand, enables us to explain why ordinary good things are good only in virtue of the fact that people are around to care about them without tempting us to the conclusion that the only good things are mental states and experiences.

To some, it may seem paradoxical to claim that things are good because we desire or choose them, rather than to say that we desire or choose them because they are good. Ross, for example, finds it clear that when we call something good we are thinking of it as having some attribute, not as an object of interest; he thinks of our interest as inspired by the perception of the thing's goodness. We choose the thing because it is good. This picture is part of what gives power to the theory that goodness is not relative to interest, and of course there is a way in which it is true. For instance, when we want a certain kind of thing, we usually want one with the virtues of that kind of thing. And it is also true that what makes a thing a good kind of thing is its virtues. In this sense our choice may be called forth by a thing's goodness, rather than the thing's being good because of our choice. But when we inquire into the basis for calling certain properties of a thing its "virtues," we always come back to something that is relative to certain conditions of human life. It is our interests and the bases of our interests that make certain qualities *virtues*; so these facts cannot make goodness a nonrelational attribute.

This shows up most clearly in the everyday kind of case of "mixed" value, in which the distinction between what we value for its own sake and what we value for the sake of something else is itself overstrained. Take this case: there are instrumental reasons, good ones, for eating. It keeps you alive. But most people could not really be said to eat in order to stay alive. Certainly, only someone who didn't enjoy eating, perhaps because of illness or some damage to the taste buds, would say that he ate "in order to stay alive." Are we then to say that eating is an activity that also has an *intrinsic* value? (Perhaps then we should be glad that we are so constituted that it is necessary for us?) Or shall we say that people eat for the sake of enjoyment – that pleasure is an end to which eating is a means? Of course, you cannot exchange another pleasure for it; hunger pains will prevent that. Perhaps then we should say we eat as a means, not to obtain life and health, but just to avoid pain. Now the philosopher wants to say: the real end is painlessness. But again, only someone in a particular situation would say that. Is this then a complicated case, to which the ends of life and health, enjoyment, and painlessness contribute in various ways? And if this is a complicated case, where are we going to find a simple one? It is easier to say that food is a good thing under the condition that you are hungry – or rather, under the set of physiological and psychological conditions that make it both necessary and pleasant for human beings to eat. Those conditions determine what the virtues of a good meal are, and not all of these virtues are instrumental properties. But this does not mean that you choose the meal in response to a perception of its intrinsic value, or of the intrinsic value of eating it. The conditions of our lives make various things valuable to us in various ways, and it is sometimes artificial to worry about whether we value those things as means or as ends. It is the conditions themselves that make the things good, that provide the various reasons

for their goodness. The question is not whether the thing possesses a special attribute, but whether these reasons are sufficient to establish the goodness of the thing.

This point can be sharpened if we distinguish between the initial condition that makes an object a candidate for choice and the full complement of conditions that, when met, renders the thing good. In the cases under discussion in this paper, the initial condition is the thing's desirability as an end (or at least not merely as a means). I have tried to show that the sense in which we can be said to desire things because they are good – i.e., for their virtues – does not show that a desirable thing need have a nonrelational property of goodness. What we call virtues just are the features of the thing that, given our constitution and situation, we find appealing or interesting or satisfying to our needs. It remains just as true, as far as this goes, to say that the thing is good because we desire it as to say that we desire it because it is good. For its virtues are still relative to our desires, or, more accurately, to the conditions that give rise to those desires. The reason that one cannot, on a Kantian account, rest with the perhaps less paradoxical formulation that value is conferred by desire is that desire is not by itself a *sufficient* condition of the goodness of its object. This is shown initially by the sort of case in which one has a desire which one would be better off without. Short of endorsing Kant's view that "the inclinations themselves, as sources of needs, however, are so lacking in absolute worth that the universal wish of every rational being must be to free himself completely from them" (G 46/428), we can agree that there are desires that conflict with one's health or happiness or that are self-destructive or pathological or simply burdensome out of all proportion to any gratification their fulfillment can provide. This already shows that the existence of a desire is not by itself a sufficient reason for the realization of its object; further conditions exist. The criterion that reasons be universalizable will also, on Kant's account, limit the capacity of desires to serve as reasons and so to confer value. But although desirability is not a sufficient condition of goodness, it is still the initial condition of the goodness of many good things, and so a main source of the goodness of those things.[19] On the Kantian view, not everything valued as an end need be intrinsically valuable or self-justifying for there to be a sufficient reason for it. A conditionally valuable thing can still be fully justified, if the unconditioned condition of its goodness is met. Things that are not self-justifying can be justified by something else. In particular, ends whose condition is their desirability can be justified by the rational choices of human beings.

6.

But I have not meant to suggest that Moore himself is prey to all of the difficulties that arise when the two distinctions are conflated. Moore has his own way of dealing

[19] I would like to thank the Editors of the *Philosophical Review* for prompting me to clarify the roles of desire and choice in conferring value.

with these issues of "mixed" value, a problem in which he was keenly interested. In order to handle cases of mixed value, Moore introduced a device which he regarded as one of his best discoveries: the theory of organic unities. The theory of organic unities involves two important points. First, it turns out that intrinsic value, on Moore's account, usually belongs to "organic" wholes or complexes of certain kinds, not to simple things. Second, it is true of such a complex whole that its value "bears no regular proportion to the sum of the values of its parts":

> It is certain that a good thing may exist in such a relation to another good thing that the value of the whole thus formed is immensely greater than the sum of the values of the two good things. It is certain that a whole formed of a good thing and an indifferent thing may have immensely greater value than that good thing itself possesses. It is certain that two bad things or a bad thing and an indifferent thing may form a whole much worse than the sum of badness of its parts. And it seems as if indifferent things may also be the sole constituents of a whole which has great value, either positive or negative.[20]

In his last chapter, "The Ideal," Moore provides various examples. For instance: the mere existence of what is beautiful has *some* intrinsic value, but so little as to be negligible, compared to the consciousness of beauty. If the consciousness of beauty is taken to be the cognition of beauty, then it in turn is made much more valuable if accompanied by an appropriate emotional response, which Moore identifies with the appreciation. Yet appreciation of beauty is not an end to which the beautiful object is a mere means. If this were so it would not matter whether the appreciation were produced in us by something genuinely beautiful or not, and it does: appreciating something that is ugly may be bad. Instead of saying that the value of the appreciation is conditional upon its appropriateness, as one might expect, Moore says that the great intrinsic value of appreciating beauty does not belong either to the object or the appreciative state but only to the complex whole formed of both. But the goodness of the whole is not the sum of the value of the parts. For we have seen that the value of the appreciation, in another context, can be absolutely negative. Moore has similar things to say about his other cases: for instance love itself is a good thing, but if your beloved is a good person, the whole is better by *more* than the addition of your beloved's goodness. These conclusions are arrived at by the method of isolation: we compare the value of various isolated wholes, with and without the relevant element. The important thing is to avoid the mistake of thinking that the element itself possesses all of the value of the difference its presence makes. It was because of this mistake that the Greeks attributed intrinsic value to knowledge. Moore explains that, really, knowledge by itself has little or no value, but that it "is an absolutely essential constituent in the highest goods, and contributes immensely to their value."[21] Similarly, the great value that has been placed upon pleasure, and the delusion that pleasure is the sole good, is attributed to the fact that:

---

[20] *Principia Ethica*, pp. 27-28.
[21] *Ibid.*, p. 199.

> Pleasure does seem to be a necessary constituent of most valuable wholes; and since the other constituents, into which we may analyse them, may easily seem not to have any value, it is natural to suppose that all the value belongs to pleasure.[22]

Indeed, getting the right account of the relation of pleasure to other sorts of value seems to have been one of Moore's major motives in introducing the idea of organic unities. Things like pleasure and knowledge have what the tradition has called "contributive value."

I hope it is evident from these examples that the principle of organic unities is meant to do the same job that the notion of a conditioned good in Kant's theory does: it allows us to say of certain things that they are valuable only under certain circumstances, or valuable only when certain other things are true or present, without forcing us to say that these kinds of things must be valuable merely as instruments. Contributive value takes on the role that conditional value plays in Kant's view. The remaining difference is that Moore makes no distinction between what would be in Kant's terms really unconditionally (or intrinsically) good and what would be objectively good as an end.

7.

The principle of organic unities is crucial to Moore, for it enables him to make some of the same distinctions and judgments that the Kantian divisions make possible. Like the Kantian distinctions, it gives us a more flexible way of talking about the value of everyday things; and like the Kantian distinctions, it makes it possible for us to explain the conditional character of a good thing without rendering that good thing a mere means. Moore, who separates pleasure from the consciousness of pleasure, even complains in one passage that if pleasure were the sole intrinsic good, consciousness would have to be regarded as a means to it.[23] But the principle of organic unities is also in a certain way perverse. The seeming difficulties that it solves in fact arise from the relational or conditional character of the value of most of the things that human beings regard as good. Yet it is precisely this relational character that Moore, with his insistence on intrinsic value, wants to deny.

Suppose someone said: on Kant's view happiness is a conditioned or extrinsic value and the good will its condition. But the happiness of a good person is, on Kant's view, always good, good under any and all circumstances, for its condition is met. So couldn't we say of this, as well as of the good will, that it is intrinsically valuable? What this would amount to would be constructing an organic unity out of happiness and the good will, and showing that on Moore's account it has intrinsic

[22] *Ibid.*, p. 93.
[23] *Ibid.*, p. 93.

value. Then the Kantian notion of "objective value" and Moore's "intrinsic value" are not so different after all.[24]

And the answer to that is that there is still a difference. For Moore's view, and the intuitionistic method of isolation, veil or obscure the internal relations within the organic unity in virtue of which the organic unity has its value. Whereas the Kantian account, which focuses on rather than ignoring the internal relations of the valuable whole, allows us to see why happiness is valuable in just this case and not in another case. Moore can only say that the combination of happiness and good will works (is a good recipe, so to speak) while happiness plus the bad will does not. Kant can say that happiness in the one case is good because the condition under which it is fully justified has been met (roughly, because its having been decently pursued makes it deserved). Those internal relations reveal the *reasons* for our views about what is valuable, while Moore's view tends to cover up these reasons. And this might be true in other cases as well: if we think that aesthetic response is only valuable when the object responded to is genuinely beautiful, or that friendship is only valuable when your friend is good, or even if we think that aesthetic response and friendship are just *more* valuable in these cases, then this has something to do with the reasons we think these kinds of things are valuable at all. On Moore's account the only relation in which the elements of an organic whole stand to each other is the relation of being elements in a single organic whole. They are all on a footing with one another. But if Kant is right there is an order within "valuable wholes," a conditioning of some elements by others, that is hidden by treating these elements as just so many *ingredients.* This order reflects the reason why the wholes are good.

Another way to put the point is this: Moore's theory drives a wedge between the reason why we care about something and the reason why it is good. Or rather, since on Moore's theory it is a mistake to talk about why something is good, we should say that it drives a wedge between our natural interest in something and our moral interest in it. On Moore's theory if you say that the reason something is good is because someone cares about it, that could only mean that the person's interest was an element in an organic whole which had intrinsic value. But according to Moore the question why such a whole has intrinsic value must not be raised: it just has the property of intrinsic value; there is no reason why it has that property.[25] Yet it is because it has intrinsic value that we ought to make it an end in our actions. A thing's goodness becomes a property that we intuit and respond to in a way that seems curiously divorced from our natural interests.

The interesting thing about that is that Moore took up the idea of intrinsic value because he saw that objectivity was not all that we wanted from a theory of value. He was certainly right to think that the same people who are discouraged by subjectivism are discouraged by an evolutionary theory or others of that kind. But to me it seems that this discouragement has to do with the way in which such theories

---

[24] Ross does something very like this in his discussion of the relation of virtue and pleasure in *The Right and the Good*, pp. 135 ff.

[25] See *Principia Ethica*, pp. 142-44.

undermine the nature of our concern for the good. For instance, if goodness is mere fitness to survive, then the only way goodness matters is the way the biological survival of the species matters – and that doesn't cover everything we feel about the importance of living a good life. But what is the nature of our concern for intrinsic values as Moore describes them? Moore seems to find it obvious that when we have determined what is intrinsically good we shall have an interest in bringing that into the world. His anti-naturalistic arguments prevent him from giving any account of why this nonnatural property should be so appealing to us. Of course, the isolation test by which intrinsic values are discerned guarantees that we will only attribute them to something that appeals to us. But that does not provide a justification of our interest in the intrinsically valuable or even a motivational explanation of it. On the Kantian account, by contrast, the good end is the object of a rational choice. The things that we want, need, care for, are good so long as certain conditions of rational choice are met. Thus, the reasons that things are good bear a definite relation to the reasons we have for caring about them.

The primary advantage of the Kantian theory of goodness is that it gives an account of the "objectivity" of goodness that does not involve assigning some sort of property to all good things. Good things are good in the way that Ross describes as relational, because of attitudes taken up towards them or because of other physical or psychological conditions that make them important to us. Only one thing – the good will itself – is assigned an intrinsic value or inner worth, and even the argument for that is not ontological. If we regard ourselves as having the power to justify our ends, the argument says, we must regard ourselves as having an inner worth – and we must treat others who can also place value on their ends in virtue of their humanity as having the same inner worth.

If human beings have an intrinsic value by virtue of the capacity for valuing things, then human beings bring goodness into the world. The distinction between a thing that is intrinsically good and a thing that is extrinsically good yet valuable as an end allows for the possibility that the things that are important to us have an objective value, yet have that value because they are important to us. Objective goodness is not a mysterious ontological attribute. The things that are important to us can be good: good because of our desires and interests and loves and because of the physiological, psychological, economic, historical, symbolic and other conditions under which human beings live.

CHAPTER 9

S. KAGAN

RETHINKING INTRINSIC VALUE

The phrase "intrinsic value" is something of a philosophical term of art. It is not, I think, an expression in much use in ordinary conversation or writing, not even among the fairly educated. Yet, at the same time, it also seems to me plausible to suggest that when philosophers introduce the term "intrinsic value" they are attempting to provide a label for a concept that does occur in ordinary thought, even if it only occurs implicitly and without a common label.

As an analogy, think of the concept of supervenience. Although one could hardly suggest that the term "supervenience" is anything other than a bit of philosopher's jargon, I think it clear that it is meant to provide a label for a certain kind of dependence relation the basic idea of which is indeed to be found (even if only implicitly) in ordinary thought. That is to say, I think that many people have the *concept* of supervenience, even if they don't have the *term* supervenience – indeed, even if they don't have any term for the concept in question at all. Philosophers provide a label for that dependence relation, and this then allows all of us to theorize explicitly about it.

Similarly, then, it seems to me that our first bit of philosophical jargon, "intrinsic value," is also meant to name a concept that most people do have – the concept of a particular kind of value – even if they don't have this or any other term for the concept in question. Providing the label allows us to theorize explicitly about that type of value.

I say all this not so much to make a fetish out of ordinary beliefs about the concept of intrinsic value, but rather to warn us against the opposite danger, that is, that we will make a fetish out of philosophical beliefs about the concept of intrinsic value. In particular, it seems to me that the very label we have provided ourselves – "intrinsic value" – reflects a philosophical theory about the nature of the value in question. And it seems to me that this theory may well be false.

Nonetheless, there is a strong temptation to think that the philosophical theory in question *must* be true. Realizing that the phrase "intrinsic value" is indeed a philosophical term of art, and given that the theory is effectively "built in" to the

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 97-114.
© 2005 *Springer. Printed in the Netherlands.*

term itself, there is a strong temptation to think that the theory must be true, by definition. It is this temptation that I am especially keen to resist.

Let me start, then, by distinguishing two concepts, both of which have some claim to being considered concepts of intrinsic value. On the one hand, we have the notion of the value that an object has independently of all other objects – the value that an object has "in itself." Philosophers sometimes try to get at this kind of value by suggesting that it is the value that an object would have even if it were the only thing existing in the universe. Although this particular suggestion is not without its own difficulties, it points us toward the basic idea that value of this sort must depend solely upon the intrinsic – that is, roughly, nonrelational – properties of the object.[1] After all, if the object's intrinsic value is had independently of all other objects, that value cannot depend at all upon any of the relational properties of the object; rather, its intrinsic value must depend upon the intrinsic properties of the object alone. It is, of course, a further question whether anything at all does have intrinsic value in this first sense. But such value, if it does exist, depends on an object's intrinsic properties alone.

This first notion of intrinsic value should be distinguished from a second concept, that of the value that an object has "as an end." I suppose that the familiar picture at work here goes something like this. Many objects are valued merely as means to other objects – they are valuable solely by virtue of the fact that they will produce (or help produce) those other objects. Those things valued as a means in this way possess "instrumental" value. But what about the objects that the instrumentally valuable objects are means to? In some cases, of course, objects may possess instrumental value by virtue of being means to objects that are themselves of no more than instrumental value (as means to still other objects). But eventually – or so the thought goes – we must reach objects that are valuable as "ends" or "for their own sake." The objects that come at the *end* of these chains – those that are desired (or deserve to be desired) for their own sake – have intrinsic value in the second sense of the term.

I think this familiar picture is itself problematic, for reasons that I will get into later. But its very familiarity should help to fix this second notion of intrinsic value, value as an end.

It is, once again, a further question whether there is anything at all that does have value as an end (although I imagine that very few will deny that there is). But if something does have value as an end, then there is reason to "promote" it, to try to

---

[1] The contrast between intrinsic and relational properties is indeed only a rough one, since (as Gary Rosenkrantz pointed out to me) some relational properties are actually intrinsic (for example, certain relations between an object and its parts). For simplicity, however, I'll continue to refer to "relational properties"; this can be read as shorthand for "*nonintrinsic* relational properties." Obviously, giving an adequate characterization of the distinction between intrinsic and nonintrinsic properties would be a difficult matter; I won't attempt that here.

produce the valuable object, or perhaps to preserve and maintain it; we sometimes say that the world is better off "as such" for the existence of the valuable object.[2]

In laying out these two notions of intrinsic value I have helped myself to various phrases. I've spoken of something being valuable "in itself," or "as an end," or "for its own sake." And I've spoken of the world being better off "as such." Let me hasten to admit that none of these phrases wear their meanings on their sleeves, and not everyone will feel altogether comfortable with the uses to which I have put them, aligning them with one or another of our two notions of intrinsic value. The same would no doubt be true for various other phrases that sometimes get used in similar discussions (for example, talk of something being valuable "in and of itself"). I doubt that any of these terms unambiguously pick out a single one of our two concepts, and so my own uses are somewhat stipulative. I can only hope, however, that I have said enough to give a rough feel for the two concepts that I have in mind.

For it does seem to me clear that these are indeed two distinct concepts. And on the face of it there is no reason to assume – at least without further argument – that the two kinds of intrinsic value come to the same thing. That is, it seems to me to be a substantive claim that whatever has value as an end has this value solely by virtue of its intrinsic properties.

Of course, one could accept this substantive claim, and I think that the dominant philosophical tradition here does exactly that. Indeed, as I have already suggested, I think that the dominant philosophical tradition finds expression in the very term that has been introduced to facilitate the discussion – "intrinsic value."

More particularly still, what I find plausible is this: the concept of value as an end is one that plays a role – even if unlabeled and often only implicit – in ordinary thought. Seeking to gain the genuine advantages of explicit investigation, philosophers have introduced a label for this notion of value as an end. But accepting the substantive thesis that an object's value as an end depends solely upon its intrinsic properties, they have settled upon the label "intrinsic value" – thus enshrining that thesis into the very vocabulary with which we discuss value.

Now to be honest, I am not a historian, and I don't know whether or not my speculative reconstruction of intellectual history is accurate. Indeed, to be utterly honest, I don't even care whether or not my bit of historical story-telling is accurate. What I am keen to argue, however, is that it is indeed a substantive thesis that value as an end depends solely upon intrinsic properties. If we don't see this as the substantive thesis that it is, this is (I suspect) because we use the single, theory-ladened label to pick out both concepts. But whatever the cause of our unthinking allegiance to the thesis, we do well to rid ourselves of it. I want to argue that the

[2] There are, of course, still other questions as well. For example, does everyone have reason to promote every object that has value as an end, or is such value, rather, "agent-relative," so that for any given valuable object only particular individuals have reason to promote it? But I will leave these important questions aside.

substantive thesis is false. Or at least (a bit more cautiously) I want to argue that we should not assume the thesis to be true without considerable argument.

That is to say: it seems to me fairly *likely* that value as an end need *not* depend solely upon an object's intrinsic properties. But even if I am wrong about this, I am convinced that the substantive thesis should not be accepted without argument – argument that to date has not been forthcoming.

Since I intend to continue to refer to value as an end as intrinsic value, my own favored way of stating my particular thesis is that intrinsic value need not depend solely upon intrinsic properties. To those trained in philosophy, however, this claim can appear incoherent. But of course, I am not claiming that the value of an object which depends solely upon its intrinsic properties need not depend solely upon its intrinsic properties. This is indeed a trivially false suggestion. I am claiming, rather, that value *as an end* need not depend solely upon an object's intrinsic properties. This thesis may be false as well, but at the very least I hope to show that it is not *obviously* false.

If intrinsic value need not depend solely upon intrinsic properties then it may also depend at least in part upon relational properties. I'm going to offer several examples to show the reasonableness of allowing for this possibility. My aim, I should note, is not to argue for the plausibility of any given example. (Indeed, I don't myself share the views being discussed in all the examples.) But I hope the reader will agree that each of these examples presents an intelligible perspective – a position concerning the value of the given object that can be readily understood and is not without its own appeal. Even if we reject these positions after further reflection, we should not be disposed to simply rule them out of court as incoherent. We should leave conceptual space for views of these sorts. But to do this, we have to leave space for the possibility that intrinsic value may depend upon relational properties.

Ultimately, in fact, I want to go even further. It seems to me intelligible to suggest that among the relational properties that are relevant to intrinsic value are the causal properties of an object. And, in particular, it seems to me that among the relevant causal properties might be the very fact that an object has produced (or is a means to) another valuable object. Thus, I want to leave open the possibility that the *intrinsic* value of an object may be based (in part) on its *instrumental* value.

If I am right about this, of course, then the familiar contrast between intrinsic value and instrumental value is mistaken, or at least dangerously misleading.[3] But this is a point to which we shall return. First, some examples.

1. Consider, first, a radical subjectivist, who holds that absolutely nothing would have any value as an end, in the absence of some creature who values it. Of course, given that there *are* creatures that value objects as ends, some things do have

[3] Failure to see this point seems to me the most significant error in Christine Korsgaard's otherwise commendable "Two Distinctions in Goodness" (reprinted in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 249-274 [* pp. 77-96 of this volume]). Korsgaard makes many points similar to those I make here.

intrinsic value. That is, the subjectivist does believe that many objects do indeed possess value as an end. Obviously, however, they do not possess that value solely by virtue of their intrinsic properties. For it is not an intrinsic property of an object that it is valued by some creature. It is, rather, a relational property.

Adherents of the dominant philosophical tradition typically say that according to radical subjectivism nothing at all has intrinsic value. Insofar as this merely means that, according to subjectivism, nothing has value solely by virtue of its intrinsic properties, then that may be (virtually) correct.[4] But it would be dangerously misleading, at the very least, to come away from such a remark thinking that according to subjectivism nothing has value as an end – for this is far from the case. It seems to me preferable to allow for talk of intrinsic value under subjectivism, simply noting that, according to subjectivism, intrinsic value depends upon a relational property.

I might note, if only in passing, that even subjectivists are not altogether free of the temptations inherent in the dominant philosophical tradition (even though they themselves provide an important counterexample to it). For many subjectivists will want to say something like the following: for an object to be valuable as an end, not all forms of being valued are sufficient. In particular, if an object is valued merely as a means to something else, then that object possesses merely instrumental value. In contrast, for an object to be valuable as an end, it must itself be valued "intrinsically." And typically, I think, this is taken to mean that the object must be valued simply by virtue of what the object is "in itself" – independently of other objects. That is, the only objects with intrinsic value are those objects that are valued simply by virtue of their intrinsic properties.

In short, even though subjectivism constitutes a counterexample (if true) to the thesis that all intrinsic value is based solely upon intrinsic properties, many subjectivists nonetheless hold a corresponding thesis – that all intrinsic valuing is based solely upon intrinsic properties. Presumably, however, if we can free ourselves of our unthinking allegiance to the first thesis, we ought to be able to free ourselves of the second thesis as well. We should allow for the possibility that someone might value an object intrinsically – that is, as an end – even though what they value about the object is not simply a matter of its intrinsic properties. It is important to free ourselves of *both* of these theses, and indeed the best evidence that it is intelligible that intrinsic value need not depend solely upon intrinsic properties lies in the very fact that many of us do indeed *value* things intrinsically without doing so on the basis of their intrinsic properties alone.

2. Subjectivism, if true, provides a very general objection to the claim that intrinsic value depends solely upon intrinsic properties. If it is correct, then in (virtually) no case at all does the intrinsic value of an object depend solely upon its intrinsic properties. But of course one might reject this general claim – holding that in many

---

[4] Though the case of a creature that values itself may provide an exception to this generalization: isn't it an *intrinsic* property of the creature that it values itself?

cases intrinsic value does depend solely upon intrinsic properties – while still thinking that in at least some particular cases relational properties matter as well.

Consider, for example, the importance of uniqueness. Many people, I think, are attracted to a view according to which the intrinsic value of an object depends in part on how rare that object is, or (in the limiting case) on its being completely unique. Obviously enough, however, uniqueness is not a property that an object has independently of whatever else may exist in the world; it is a relational property, rather than being an intrinsic one. Thus if an object's value as an end can depend upon its uniqueness, intrinsic value need not depend solely upon intrinsic properties.

There are, of course, familiar moves that defenders of the dominant philosophical tradition can make at this point. They might suggest, for example, that uniqueness contributes at best to something's instrumental value, perhaps providing a source of pleasure, say, that cannot otherwise be attained. But the fact remains that many of us would not find it plausible to insist that an account along these lines tells the complete story. We might want to insist, for example, that beautiful objects are *intrinsically* valuable – not merely as a means, but as an end. (They are worth having "for their own sake"; the world is the richer "as such" for their existence.) And it might be, as well, that a *unique* work of art is made all the more valuable as an end by virtue of that very fact.

To forestall misunderstanding, let me hasten to point out that an advocate of this view need not think that uniqueness contributes to the intrinsic value of an object *regardless* of the other properties of the object. That is, one need not hold that anything at all becomes intrinsically valuable as it becomes rare or unique. It might be, for example, that only objects that are independently intrinsically valuable are such as to have their value enhanced by uniqueness. But even if so, the fact would remain that for such objects uniqueness would *increase* the object's intrinsic value. Hence intrinsic value would indeed depend (if only in part) on a nonintrinsic property.

Once again, the present point is not so much to evaluate this proposal as to see that it is *sufficiently* intelligible and plausible that we should want to leave room for it in our conceptual framework. But to do so, we must allow for the possibility that intrinsic value can depend on relational properties.

3. Uniqueness, of course, is not a causal property; but it does not seem difficult to think of examples where someone might well want to ascribe intrinsic value on the basis of properties that are causal. Consider an elegantly designed racing car, one capable of driving at extraordinary speeds while still handling with ease. Someone might value the existence of such a car, and indeed value it intrinsically, as an end. Thus, they might think the world is better off ("as such") for the existence of such a car; they might think they have reason to bring such a car into existence, or to preserve it and care for it.

Now if we ask what it is about the car that makes it valuable in this way, the answer will presumably make reference to its causal properties – let's say, its ability to perform at a particular speed. Thus, the suggestion being made is that the car is

intrinsically valuable by virtue of its relational properties, including its causal properties.

Once again, friends of the dominant philosophical tradition might suggest that the car is simply valuable instrumentally – as a means of going fast, or as a way of winning the race, or simply as a method of reminding ourselves of our engineering prowess.

But we need not find such alternative accounts compelling. The car itself might never have been driven at all, and indeed we might never intend to drive it. What we might find valuable is simply its *ability* to perform at the given speed. We might value its *suitability* for the task of racing. This is, of course, a causal, instrumental property, at least if we construe these notions broadly (that is, after all, my very point), but it seems strained to insist that this causal property could not conceivably be the basis of the car's being valuable as an end. Rather, the most natural way to understand the view being discussed here is as claiming that the car has intrinsic value *by virtue* of its causal properties. And whether or not you are attracted to this view, you may at least find it sufficiently intelligible that you want to leave room for its possibility.

4. Somewhat similar cases involve excellence in various practical arts – for example, fine cooking. No doubt most of us do value the ability to cook a gourmet meal at least in part for merely instrumental reasons. (Presumably, for example, we might value the ability as a means to fine food, and the food as a means to pleasure.) But I think it is not an uncommon view to hold that such abilities are intrinsically valuable as well – that they are valuable as an end, and not merely as a means.

Now it might be suggested, reasonably enough, that insofar as we do value such skills intrinsically, it is by virtue of their being manifestations of excellence; and perhaps – though this might be more of a stretch – it could also be argued that being a manifestation of excellence is itself solely a matter of the intrinsic properties of the ability. But whatever the merits of these claims, it seems to me that something more needs to be said as well, for we do not typically value excellence in any skill whatsoever – however pointless and useless the skill – even if such excellence is difficult to attain. In the case of cooking, for example, it seems to me that an important part of the reason we value the skill lies precisely in the fact that this skill is *useful.*

I hope I am not being misunderstood. I am not making the unremarkable observation that the ability to prepare gourmet food is instrumentally valuable by virtue of its being useful. Rather, I am suggesting that if we do value this ability *intrinsically*, part of the reason that we do so lies in the fact that this ability is useful. That is to say, it is the usefulness – the instrumental value – of culinary skill that provides part of the basis of the intrinsic value of that skill. Were culinary expertise to somehow lose its instrumental value (if we no longer needed food, and if it no longer gave us pleasure), it would lose at least some (and perhaps all) of its *intrinsic* value as well. Indeed, it might be suggested that something very much like this has

gone on for other practical skills, where technology has robbed a skill of its instrumental value, and thereby reduced or eliminated its intrinsic value as well.

As before, the question is not whether on reflection you will accept this view (though I think that, suitably refined,[5] it may have much to commend it), but only whether a view like this seems sufficiently intelligible that we should try to leave conceptual space for it. For if we are to do this, we will have to allow for the possibility that intrinsic value can depend, in part, on instrumental value.

5. Although the last two examples suggest that intrinsic value may depend, at least in part, on instrumental value, it should be noted that they are only cases in which instrumental *capacities* are relevant to intrinsic value. It is because of what the race car *can* do that it has intrinsic value, whether or not it ever is used. Similarly, perhaps, it is because of what culinary skill *can* do that it is intrinsically valuable. (Even if the gourmet food produced is never consumed, one might still think the skill itself is intrinsically valuable.)

So let's consider one more example, one where the actual causal history of the object is taken to be relevant to its intrinsic value. Consider the pen used by Abraham Lincoln to sign the Emancipation Proclamation, freeing the slaves. Clearly, this pen had considerable instrumental value – it was the actual means by which a great deal of intrinsic good was brought into the world. But it seems to me that we might want to say something more than this. It seems to me that we might want to suggest that this pen has *intrinsic* value – that the continued existence of this pen has value as an end. Of course, the pen's defining instrumental moment is now long since over. But by virtue of that history, we might say, it now possesses intrinsic value: it is something we could reasonably value for its own sake. The world is the richer for the existence of the pen; its destruction would diminish the value of the world as such.

As usual, no doubt, one can point to continuing instrumental value in the pen: perhaps, for example, its display in a museum edifies us, reminding us of the value of human freedom. But to insist that such ongoing instrumental value is the sole source of whatever value the pen now has seems to me to be an overly narrow view. At the very least, I think I understand someone who suggests that the pen *itself* has intrinsic value – that it is valuable as an end.

Of course, if it does possess intrinsic value, this is by virtue of the fact that this very pen played a historically important causal role. It is not the mere *capacity* to have played this role that singles out the pen as having intrinsic value: any of a large number of other pens near Lincoln could have done just as well. It is, rather, the fact that this particular pen actually played the particular causal role that it did. That is, in this case, at least, intrinsic value seems to depend not upon mere instrumental capacity, but rather upon actual instrumental history.

---

[5] To mention just one aspect in need of further specification, a view like this will presumably have to appeal to some notion of a minimal relevant level of usefulness. After all, even "useless" skills have *some* instrumental value.

Note, further, that it seems plausible to suggest that if this pen does indeed have any intrinsic value, most or all of it is due to this instrumental role. Stripped of its instrumental history, the pen probably has no intrinsic value at all. Thus, in this case, it might be suggested, the intrinsic value of an object depends *completely* upon its instrumental value.[6]

Let's pause for a moment and think about the implications of what I have just said. It is, of course, a familiar point that a single object might be both instrumentally valuable *and* intrinsically valuable. But I am arguing for something considerably stronger: I am arguing that something may have intrinsic value – in part, or even in whole – because of its instrumental value. For such objects, then, when we specify the properties by virtue of which the object possesses intrinsic value, we will need to list the instrumental properties as well.

This last point, I think, is likely to encounter particular resistance. Even some of those sympathetic to my suggestion that intrinsic value need not depend solely upon intrinsic properties may hesitate before allowing for this possibility. The objection seems obvious: insofar as X is instrumentally valuable, it is only valuable because it is a means to something *else*, Y. To *say* that it is instrumentally valuable is just to *say* that it is valuable (in that regard) merely as a means to something else. So it cannot by virtue of that very fact be *intrinsically* valuable; it cannot thereby be worth having for its *own* sake.

But it seems to me that this objection simply begs the question. To see this, let's start with a simple point. It is, of course, true that to point out that something is instrumentally valuable is indeed to say something important about *why* it is valuable (namely, that it stands in a certain causal relation to something else of value). But obviously enough, we don't want to hold that if we explain the basis of something's value, it can't possibly be *intrinsic* value that we are explaining. That is, we don't want to insist that intrinsic value is inexplicable. On the contrary, typically at least, we can explain something having intrinsic value by noting the various properties that provide the basis of its intrinsic value. Accordingly, if there really is an objection to the suggestion that in some cases instrumental value may be the basis of intrinsic value, this cannot be on the general ground that intrinsic value has no basis (and so, *a fortiori*, the basis cannot be something's instrumental value). Rather, the objection will have to be to the *particular* basis offered. That is, the objection must be that *instrumental* value can never be the basis of intrinsic value.

But to assert this, without argument, is simply to beg the question at issue. Obviously, if we could assume, without further ado, that intrinsic value must be based solely upon *intrinsic* properties, then it would indeed follow trivially that instrumental value cannot be the basis of intrinsic value. But of course this is the very assumption I am trying to challenge. And once we have allowed for the

---

[6] That is, but for its instrumental value, the pen would have no intrinsic value at all. Note, however, that to say this is not to claim that the pen's being instrumentally valuable is the *only* feature relevant to its having intrinsic value. One could still insist, plausibly, that other properties help ground the pen's intrinsic value as well. Thus, instrumental value may here be *necessary* for intrinsic value, without being sufficient.

possibility that relational properties may be relevant to something's intrinsic value, then I don't see how we can rule out the possibility that instrumental value may be among the relational properties that are relevant. At a minimum, it will take a further argument to rule this possibility out; it won't do to simply deny it without reason.

Of course, none of this should be taken to mean that I am suggesting that all cases of instrumental value ground intrinsic value. That would be a further claim, and an extremely bold one; and I see nothing to recommend it. It simply seems to me that in *some* cases instrumental value may ground intrinsic value.

If this is indeed possible, then we will need to distinguish between two types of cases involving instrumental value. In the typical case, presumably, something's instrumental value will not contribute to its intrinsic value at all. If the object in question does happen to have intrinsic value (which it need not), this will not be explained – even in part – in terms of its having the instrumental value that it does. In such cases we can speak of the instrumental value as being "mere" instrumental value. (Though we should note that an object's having mere instrumental value is quite compatible with its *also* having intrinsic value, on independent grounds. Thus, having "mere" instrumental value should not be confused with *merely* having instrumental value – that is, having instrumental value alone, and no other value, including intrinsic value.)

In other cases, however, an object's instrumental value will contribute to (and perhaps, in some cases, even be a ground of all of) the object's intrinsic value. In such cases we can speak of "intrinsically valuable instrumental value." And we can say that the object is intrinsically valuable (at least in part) *because of*, or *by virtue of*, its instrumental value.[7]

To put the point still another way, when an object has *instrumental* value, it is worth having for the sake of something else. And often that is the end of the matter. But in at least some cases, by virtue of the very fact that the object is (or was, or will be) worth having for the sake of something *else,* it is also worth having for its *own* sake as well.

6. Before moving on to other issues, I want to quickly mention one last example. What is the meaning, or value, of life? A very common answer, I believe, suggests that it lies in helping other people attain meaningful lives themselves. From the standpoint of the dominant philosophical tradition, however, this answer is deeply flawed. First of all, helping others is an instrumental relation, and so can provide at best instrumental value to one's own life. Furthermore, unless there is something *else* that can provide a life with *intrinsic* value, helping another won't even provide one's own life with *instrumental* value. So this common answer cannot be complete.

[7] Let me note, if only in passing, that similar locutions may be necessary for other types of value as well. I take it, for example, though I won't argue the point here, that symbolic value is not necessarily a form of instrumental value. And typically, no doubt, having symbolic value does not itself ground an object's having intrinsic value. But it does seem possible to me that for at least some symbols the symbolic value *does* itself provide (at least part of) the basis of the object's intrinsic value. So we may need to distinguish between "mere" symbolic value, on the one hand, and "intrinsically valuable symbolic value," on the other.

(It won't solve the second problem to suggest that my helping you has instrumental value by virtue of the instrumental value to be found in the fact that you help someone else. For chains of instrumental value – however long – must end in one or more things with intrinsic value.)

Suppose, however, that helping another is not a case of "mere" instrumental value, but rather a case of intrinsically valuable instrumental value. Then the problem is solved. In helping someone else, my own life has intrinsic value – by virtue of this instrumental fact about me. And my acts of helping another will indeed possess instrumental value, provided that I help someone who helps another (who herself helps another, and so on). My acts will possess instrumental value because they help produce lives with intrinsic value – lives that themselves possess intrinsic value by virtue of their instrumental value.[8]

In arguing for the intelligibility of a view of this sort, I do not mean to suggest that one who holds a view like this could not, or should not, admit that there are indeed various other sources of intrinsic value in our lives. I simply want to note that a view of the kind I have just sketched *could* be complete in itself – appearances to the contrary notwithstanding – once the dominant philosophical tradition is rejected.

I have been arguing, by means of a series of examples, for leaving open the possibility of an object's intrinsic value depending upon some of its nonintrinsic properties. As I have already explained, however, it is not my purpose to defend any of the specific views that lie behind the various examples. You may find, in fact, that you don't accept any of them. If so, you certainly won't be convinced that intrinsic value ever actually does turn on nonintrinsic properties. (At least, you won't be convinced on the basis of any of my examples; you might, of course, have your own.) But even if this does describe your situation, I hope that you agree that at least some of the various examples are sufficiently intelligible that we should be willing to leave open the *possibility* that relational properties may play a role in determining intrinsic value – that the possibility should not simply be ruled out of court, without further argument. And if you do agree to this much, then I have accomplished what I set out to do.

There may be some, however, with a certain amount of sympathy to my basic position, who nonetheless take issue with the way in which I choose to express it. They might agree, for example, that an object's value as an end need not depend solely upon its intrinsic properties. But they might insist nonetheless that value as an end should not be *called* "intrinsic value" – that the *term* "intrinsic value" should be reserved for the value that an object has solely by virtue of its intrinsic properties. Thus, I might be correct to distinguish value as an end from intrinsic value, and I might be correct to chastise the dominant philosophical tradition for its assumption that value as an end must depend solely upon intrinsic properties; but I am wrong to

[8] There is no threat of infinite regress here. If you and I help each other, for example, then each of our lives has intrinsic value, and each of our lives has instrumental value – and each life has its intrinsic value by virtue of its having instrumental value.

report any of this by saying that *intrinsic* value need not depend solely upon intrinsic properties.

This objection can certainly be offered in a friendly spirit. The objector might go on to offer some alternative favored term for value as an end, be it "final" value,[9] or "end" value, or "basic" value, or what have you. The thought is simply this: the term "intrinsic value" does seem to wear its meaning on its sleeve – at least to the philosophically trained – and the meaning it seems to wear is not that of value as an end, but rather the value that an object has solely by virtue of its intrinsic properties. So why don't we agree to reserve the term for that kind of value? Any other use is potentially misleading; and in any event, intrinsic value – understood as the value that an object has by virtue of its intrinsic properties alone – is certainly a *kind* of value worth studying, isn't it?

I find myself inclined to reject this friendly suggestion. I certainly agree that if we do continue to use the term "intrinsic value" for value as an end, then we leave ourselves open to natural misunderstandings – in particular we leave ourselves open to the common philosophical assumption that intrinsic value must depend solely upon intrinsic properties, the very assumption I have been challenging. Nonetheless, it seems to me that other possible names (at least, those I can think of) are themselves unsatisfactory and potentially misleading in their own ways, and so this weakens at least some of the reason we might have for abandoning the tradition of referring to value as an end as "intrinsic value."

Furthermore, I am inclined to be skeptical of the claim that intrinsic value – understood as the value that an object has simply by virtue of its intrinsic properties – does pick out a particular kind of value worthy of study. Why should we think that it does?

Remember, first of all, that to pick out the value that an object has by virtue of its intrinsic properties alone is to identify a type of value on the basis of a certain type of metaphysical fact, namely, that the relevant properties are all "one-place" properties. But why should we think that this picks out a kind of value of particular interest from the perspective of *value* theory? Why should this type of value be of any more interest to us as value theorists than it would be to pick out the value that an object has on the basis of its relational properties alone? Or the value that an object has on the basis of its 17-place properties alone?

Some, I suppose, might be tempted by the claim that value based on intrinsic properties alone is a kind of value that an object has necessarily. And necessary value would, I grant, be an interesting type of value to study. (Of course, even if this were so, we would have no reason to assume that value based on intrinsic properties would be the *only* kind of necessary value.) But in any event, the tempting thought is mistaken: since intrinsic properties need not be had necessarily, value based on intrinsic properties alone need not be possessed necessarily. So what then, *is* especially interesting about value based on intrinsic properties alone? Is there anything at all that makes it especially worthy of study?

---

[9] This is Korsgaard's proposal, in "Two Distinctions in Goodness."

Note, second, that without further argument we cannot even assume that all instances of value based on intrinsic properties alone will be instances of value as an end. (That is, to put it in my preferred terminology, we cannot assume that value based on intrinsic properties alone is always a kind of intrinsic value.) Admittedly, intrinsic properties alone can never ground instrumental value, but for all that, in any given case the value that is grounded in intrinsic properties alone might be a relatively unimportant or "lesser" value. Thus, we cannot assume – in the absence of an argument – that value grounded in intrinsic properties alone is always value as an end.[10]

I have, of course, already argued for the possibility that value as an end need not be based on intrinsic properties alone. But now I am drawing our attention to a further point: in the absence of further argument, we cannot even assume that value based on intrinsic properties alone will be an *instance* of intrinsic value. So again, we must ask, what if anything makes it especially worthy of study?

Suppose we stipulate that when we talk about the value that an object has solely by virtue of its intrinsic properties, we are restricting our attention to value as an end. That is, even if some of the intrinsic properties ground some *other* kind of value, a value that is not the basis of, or a kind of, value as an end, we are simply going to disregard that aspect of the value. So now (thanks to the stipulation) we can assume that *if* there is any value based on intrinsic properties alone, this will be a kind of value as an end.

But recall that we are now leaving open the possibility that nonintrinsic properties may be relevant to value as an end as well. (After all, we are exploring what was meant to be a friendly suggestion – open to the possibility that I have been championing.) Now take an object that has value as an end. Presumably, it will have value as an end by virtue of some subset of its properties. And for all we know, the relevant subset will include both intrinsic and nonintrinsic properties. Suppose so: then we can say that the relevant properties jointly determine the object's value as an end. So far, so good.

But does it make any sense to talk of the value that the object has *solely* by virtue of its intrinsic properties? Can the value that the object has as an end be *divided* in this way – into a part due solely to intrinsic properties, and into a different part (based perhaps solely upon relational properties, or perhaps on both intrinsic and relational properties)? I don't see any good reason to assume that this must be so.

It might be, of course, that something like this *is* so – if the overall value of the object as an end is the *sum* of more particular values as an end, based on specific subsets of properties, and if there is some guarantee that intrinsic properties make an independent contribution to this sum. But what reason do we have to think that anything like this is correct?

---

[10] Consider, for example, logical goodness (the goodness that an argument has when it is a logically good – i.e., valid – argument). This is presumably a kind of value that depends solely upon the intrinsic properties of the objects that have it (that is, arguments); yet few would take it to be an instance of value as an end. (I owe this example to Fred Feldman, "Hyperventilating about Intrinsic Goodness," in this issue of *The Journal of Ethics* [* pp. 45-58 of this volume].)

Obviously, if intrinsic properties were the only properties that could be relevant to an object's value as an end, then the difficulty would be resolved: it would indeed make sense to talk of the value (as an end) that an object has solely by virtue of its intrinsic properties – for this would simply be its value as an end. But once we allow for the possibility that nonintrinsic properties can be relevant as well, there is no reason to assume – in the absence of further argument – that the contributions to value as an end made by the various relevant properties can be segregated in this way. And in particular, there is no reason to assume that it even makes sense to talk of "the" value contributed by the intrinsic properties. Rather, it may only make sense to talk about the various ways in which the intrinsic properties contribute – together with the relevant relational properties – to the object's value as end.[11]

It may be helpful to bear in mind that few contemporary philosophers, if any, would assume that each relevant intrinsic property makes its own independent contribution to an object's overall value as an end. Instead, most (or all) would allow for "interaction effects" between the relevant intrinsic properties. But once we allow for the possibility that nonintrinsic properties may be relevant to value as an end as well, there is no reason to assume (without further argument) that the intrinsic properties nonetheless together make an independent contribution to that value – independent, that is, of the relevant *nonintrinsic* properties. For all we know, we should expect interaction effects here as well. Indeed, it might be that every contribution to value as an end made by the relevant intrinsic properties depends upon the object's nonintrinsic properties as well. Talk of the value contributed by intrinsic properties *alone* appears to rest upon the undefended assumption that such systemic interaction effects won't arise.

There are, then, several reasons to be skeptical of the thought that the notion of the value that an object has based solely upon its intrinsic properties is one that is worthy of study from the perspective of value theory. Indeed, the very thought that it makes *sense* to talk about the value that an object has in this way may rest on a mistake.

And this explains my uncharitable hesitation to accept the suggestion that we reserve the term "intrinsic value" for the value an object has solely based on its intrinsic properties, and find some other term for the value something has as an end. I am not at all sure there is anything of interest that we would be saving the label for. Meanwhile, a perfectly important category – value as an end – would go lacking a familiar and evocative label.

So I am inclined to reject the friendly suggestion, and I propose instead that we reverse the proposal. I suggest that we reserve the term "intrinsic value" for value as an end, and leave it to others to come up with a short label for the value that an object has simply by virtue of its intrinsic properties – once they convince us that there is indeed some value in finding a label for this other category!

---

[11] See Shelly Kagan, "The Additive Fallacy" (*Ethics* 99 (1988), pp. 5-31), for a related discussion of a similar point from another part of ethics.

Finally, let me mention one other proposal that might be offered in a tolerably friendly spirit. I have, of course, been talking as though various kinds of *objects* are the possessors of intrinsic value. And although I have been prepared to use the idea of an object rather broadly, to include, for example, acts and lives and skills, many of my examples have been ordinary material objects – people, and cars, and pens.

It might be suggested, however, that although it is a common enough practice to view objects as the bearers of intrinsic value, it is nonetheless preferable to hold that facts (or, perhaps, states of affairs) are the only genuine bearers of intrinsic value. If a view like this is correct, there will of course be fairly easy translation from the common informal object-based idiom to the strictly correct fact-based idiom. Instead of saying that Lincoln's pen has intrinsic value, for example, by virtue of its having been used to sign the Emancipation Proclamation, we will say that what has value is the *fact* that there exists a pen which was used to sign the Emancipation Proclamation. More generally, wherever we might say that object O has intrinsic value, by virtue of having properties P, Q, and R, we will say that what has intrinsic value is the fact that there is an O that has P, Q, and R.

I won't here enter into the various arguments that might be used to support (or attack) the claim that facts are the only bearers of intrinsic value. Note the following, however: it is not implausible to suggest that it is an intrinsic property of a given fact that it concerns the specific objects and properties that it does. That is, it would not be implausible to claim that it is an intrinsic property of the *fact* that there exists a pen which was used to sign the Emancipation Proclamation – that it is about a pen, that it is about a pen having been used to sign the Emancipation Proclamation, that it is about that pen existing, and so forth. (In contrast, it is not an intrinsic property of that fact that I am now writing about it.)

If we accept this claim about the intrinsic properties of facts (and for the sake of argument, I simply propose to grant it), and we then combine it with the earlier claim that strictly speaking only facts are the bearers of intrinsic value, then the following result emerges: one can accept the thrust of all of my examples, while still accepting the dominant philosophical tradition that intrinsic value turns solely upon intrinsic properties.[12]

Consider Lincoln's pen, once again. I claimed that we should leave open the possibility that it has intrinsic value, given the particular causal role it played. Thus, it seemed, we had to leave open the possibility that nonintrinsic properties (and, in particular, instrumental properties) were relevant to intrinsic value. But if we claim, instead, that strictly speaking the only bearers of intrinsic value are facts, we will say that what has value is (let's say) the fact that there exists a pen that was used to sign the Emancipation Proclamation. We can recognize that this fact has intrinsic value by virtue of its being about a pen being used in a particular way – but since we are

[12] This point is made with regard to the particular case of pleasure in Fred Feldman, "On the Intrinsic Value of Pleasures" (*Ethics* 107 (1997), pp. 448-466). I presume that he also sees that the move can be generalized, although he doesn't say so explicitly there.

assuming that this is an intrinsic property of the fact in question, it will still be true that only intrinsic properties of the fact are relevant to its intrinsic value.

To state the point generally, let's introduce some jargon. When it is a fact that O has properties P, Q and R, let's call O the *object* of the fact and let's call P, Q, and R the properties *ascribed* by the fact. Note that typically the properties ascribed by a fact are not themselves properties *of* the fact: we are not saying that the *fact* that O has P *itself* has P.

What my various examples have in common, then, is this: I have tried to point to cases where there is some object, O, which might be thought to have intrinsic value by virtue of having various relational properties P, Q, and R. I took this to show the appropriateness of leaving open the possibility that intrinsic value might depend on relational properties. But if the only bearers of intrinsic value are facts, then it is irrelevant to note that P, Q, and R are themselves relational properties, for these are only properties ascribed by the fact, and not (in the examples I've given) properties *of* the fact. Admittedly – if my examples are to be believed – then the *fact* that O has P, Q, and R has intrinsic value, and presumably it only has intrinsic value because (let's say) it concerns the particular object that it does, and ascribes to that object the particular properties that it does. Thus, what is relevant to the intrinsic value of a given fact is the property of concerning the particular object that it does, as well as the property of ascribing to that object the particular properties that it does. But these properties are themselves intrinsic properties of the fact (or so we are assuming). So the intrinsic value of the fact turns solely upon its intrinsic properties, and the dominant philosophical tradition remains correct when it insists that intrinsic value is based on intrinsic properties alone.

I am inclined to respond to this proposal by suggesting that it preserves the letter, but not the spirit, of the dominant philosophical tradition. Admittedly, if facts are the only bearers of intrinsic value, and if it is an intrinsic property of a given fact that it concerns the particular object that it does, and that it ascribes the particular properties that it does, then it seems likely that something's intrinsic value is determined solely by its intrinsic properties. Thus is the letter of the tradition preserved.

But note that this approach places no restrictions whatsoever on what the relevant ascribed properties are like. That is, when a given fact has intrinsic value it will have that value by virtue of ascribing various specific properties to the object – and nothing at all guarantees that the ascribed properties will themselves be intrinsic properties. Thus, the fact that O has P may have intrinsic value, but only by virtue of its ascribing P to O; and P itself may well be a nonintrinsic property. And if the examples I have given are accepted, then this is exactly what we will sometimes find: the *ascribed* properties relevant to intrinsic value will not themselves always be intrinsic properties.

This, it seems to me, still represents an important departure from the dominant philosophical tradition about intrinsic value. At a minimum, I think it represents a failure to capture some of the spirit of that tradition. I suspect that most (though, no doubt, not all) friends of that tradition would want to say something like the

following: if facts are indeed the only bearers of intrinsic value, then we must also place a restriction on what sorts of *ascribed* properties are relevant to intrinsic value. Only intrinsic ascribed properties can be relevant. That is, if a fact has intrinsic value, it can only be by virtue of ascribing intrinsic properties to its object.

I take it, after all, that most friends of the tradition would insist that uniqueness cannot be relevant to intrinsic value (that the fact that something is unique cannot be intrinsically valuable). They would insist, similarly, that usefulness cannot be relevant to intrinsic value (that the fact that something is useful cannot be intrinsically valuable). And they would insist as well that instrumental value cannot be relevant to intrinsic value (that the fact that something has played a particular instrumental role cannot be intrinsically valuable). But to insist upon these things – within the framework in which only facts have intrinsic value – they must insist that the ascribed properties relevant to a fact's intrinsic value must themselves be intrinsic properties.

But it is precisely this suggestion, I think, that we should now resist. Even if we agree that only facts are bearers of intrinsic value, we should insist that there is no good reason to assume (in the absence of further argument) that the only ascribed properties relevant to intrinsic value are themselves intrinsic properties. And this, I believe, still flies in the face of the dominant philosophical tradition.

The point perhaps can be put this way: certain "base level" properties – properties that can be had by ordinary objects, among other things – are relevant to intrinsic value. If *objects* are the bearers of intrinsic value (as I have assumed for most of this paper) then these properties are "directly" relevant to intrinsic value: objects will have intrinsic value by virtue of having these properties. If, on the other hand, only *facts* are bearers of intrinsic value, then these base level properties are only "indirectly" relevant to intrinsic value, but they are, nonetheless, still relevant: a fact will have intrinsic value by virtue of ascribing these properties to its object.

The dominant philosophical tradition (or so it seems to me)[13] takes a stand concerning these base level properties, whether they are viewed as directly relevant (if we take an object-based approach) or indirectly relevant (if we take a fact-based approach): it holds that the base level properties relevant to intrinsic value must themselves be intrinsic properties. But if the examples I have discussed are to be believed – or even if we merely find some of them sufficiently intelligible that we want to leave conceptual room for cases like them – then the dominant philosophical

---

[13] But perhaps I misunderstand the dominant philosophical tradition on this matter. After all, some philosophers firmly within that tradition do note the relevance to intrinsic value of such apparently nonintrinsic properties as *knowing*, as opposed to merely believing. (I owe this objection to Ben Bradley.) Note, however, that precisely in such cases friends of the tradition typically feel the need to start talking about the intrinsic value of (facts about) complex *wholes* (consisting of, for example, the knower and the object known). This allows them to insist that what is actually relevant to intrinsic value is simply an intrinsic property *of the whole* (roughly, that one part, the knower, stands in the right relation to another part, the object known). Yet it is often difficult to see what motivates the turn to wholes in this way, except the very belief that the *only* base level properties that are relevant to intrinsic value are *intrinsic* properties.

tradition must still be rejected. The move to a fact-based approach cannot – all by itself – eliminate our need to challenge the tradition.

To be sure, freeing ourselves of the grip of what I have been calling the "dominant philosophical tradition" concerning intrinsic value will not be easy. For as I have readily conceded, that tradition is reflected in the very terminology that we use to name and discuss the kind of value with which I have been concerned – *intrinsic value.* But it is time to challenge our unthinking acceptance of that tradition. It is time, I say, to rethink intrinsic value.

CHAPTER 10

W. RABINOWICZ AND T. RØNNOW-RASMUSSEN

A DISTINCTION IN VALUE: INTRINSIC AND FOR ITS OWN SAKE[1]

Several philosophers have recently argued for a distinction between *intrinsic* and *final* value. Following G. E. Moore (1903, 1922, 1993), an object is thought to be intrinsically valuable insofar as its value depends on ("supervenes on") its internal properties. It has final value if it is valuable "as an end", "for its own sake", rather than for the sake of something else.[2] From a traditional point of view in moral philosophy, this seems to be a distinction without a difference: to value something for its own sake is to value it for the properties it has in itself, i.e., to value it intrinsically, on the basis of its internal properties.

This traditional view, however, has come under attack in recent years, notably by such philosophers as Christine Korsgaard (1983), Shelly Kagan (1992) and John O'Neill (1992). All of them argue for the following claim:

(C) There are final values that are not intrinsic.

Entailment in the other direction has nearly never been questioned, as far as we know: it seems to be a general if not a universal view that all intrinsic values are

---

[1] This paper is dedicated to Peter Gärdenfors on the occasion of his 50th birthday. The early drafts were presented in February 1999, at a Lund-Copenhagen symposium on intrinsic value, and then in March and May 1999, at the Departments of Philosophy in Stockholm and St. Andrews. We are indebted to the participants of these meetings, and in particular to Lars Bergström, John Broome, Gunnar Björnsson, Krister Bykvist, Garrett Cullity, Åsa Carlson, Johan Brännmark, Klemens Kappel, Ulrik Kihlbom, Ian Law, Hans Mathlein, Derek Parfit, Ingmar Persson, Peter Sandoe, Caj Strandberg, Folke Tersman and, last but not least, Michael Zimmerman, for their interesting comments. We have also benefited from discussions with Bengt Brülde, Erik Carlson, Sven Danielsson, Dan Egonsson, Sten Lindström and Björn Petersson.

[2] A word of warning: "value as an end" is a somewhat misleading expression in the present context, as will be argued in the last section of this paper. The "for its own sake"-idiom is more appropriate for what we have in mind.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 115-129.
© 2005 *Springer. Printed in the Netherlands.*

final.[3] In this paper we do nothing to question this view. Instead, we focus on (C) and try to give further support to this claim.[4]

Before we continue, it might be appropriate to clarify our usage of the notion of an internal property. The nowadays common identification of such properties with the non-relational features of the object should be qualified: as we understand the term, the internal properties of an object are not just its non-relational properties but also the "internally relational" properties that it possesses in virtue of the relations it has to its own *parts* (components, elements, constituents, etc.). Thus, it is an internal property of a car that it has, say, four seats. And it is an internal property of a set that it has a car among its elements.[5] For convenience, however, we shall in what follows use the expression "relational property" as short for "externally relational property" (i.e., a property that an object has in virtue of its relation to something that is not one of its parts). Given this usage, relational properties will be contrasted with the internal ones.

We should also point out that, in our usage, an internal property of an object need not be essential to it. A concrete individual, such as, say, Theseus' ship, may well have internal properties that are contingent, say, such features as its colour or the property of containing as a part a particular plank, *a*. Abstract objects are a special case: *their* internal properties are necessary to them; they are essential to their identity.[6] Note that the intrinsic value of a concrete individual object may well vary in different possible worlds, if this value supervenes on the object's contingent internal properties. Moore himself did not allow for this possibility. He insisted that

---

[3] The only exception we know of is G. E. Moore, who in his *Ethics* suggests that for a thing to be "good for its own sake" it must be intrinsically good through and through: it cannot contain parts that are intrinsically bad or indifferent. Thus, for a (hedonistic) utilitarian, only pleasure is good for its own sake, while various wholes containing pleasure along with other components can at best be intrinsically good. They can be good for the sake of the pleasure they contain but not for their own sake. (Cf. Moore, 1965 (1912), pp. 30-32.) In what follows, however, we ignore this special Moorean usage of "for its own sake".

[4] When this paper had already been completed, we came across a new paper by Kagan (1998) [* pp. 97-114 of this volume]. There are several striking similarities between his present views and ours, along with some important differences. See footnote 29 below, for a short summary of the similarities and the differences.

[5] See Moore (1922, pp. 261 f., and 1993, pp. 26 f.). Moore would in addition require that the internal properties, both the non-relational and the internally relational ones, should all be *purely qualitative* in the broad sense of not being dependent on the numerical identity of objects. The "intrinsic nature" of an object, *a*, that comprises all of its internal properties, does not contain such properties as "being *a*" or "having *b* as a part". According to Moore, a part of what is meant by the claim that the intrinsic value of a thing depends solely on its intrinsic nature is that "anything *exactly like it*" must possess this value in the same degree (Moore, 1922, p. 261; cf. Zimmerman, 2001, ch. 3). The phrases "having a different intrinsic nature" and "not exactly alike" are "equivalent" for Moore (1922, p. 261). The same idea appears in Moore's posthumously published "Preface to the second edition" of *Principia Ethica*, where he interprets intrinsic properties as properties that are invariant under the relation of being exactly alike (Moore, 1993, p. 23; cf, also ibid, p. 27). It seems to us that this requirement of "qualitativeness" should be avoided. That intrinsic value only depends on purely qualitative properties is certainly true; it is an implication of the universalizability principle. But it should not be made true just by the definition of an intrinsic value.

[6] For a short discussion of these issues, see Rabinowicz and Österberg (1996, p. 23). Cf. also Zimmerman (2001), ch. 3, who refers to Humberstone (1996) for an extended discussion of the different interpretations of the notion of an intrinsic (internal) property.

"it is *impossible* for what is strictly *one and the same* thing to possess that value at one time, or in one set of circumstances, and *not* possess it at another" (Moore, 1922, p. 260; cf. also Moore, 1903, p. 78, and 1993, p. 22). This seems to imply that the internal properties of a thing, on which its intrinsic value supervenes, must all be necessary to the thing in question. Here, however, we have decided to allow for contingent internal properties. This choice squares better with at least some of Moore's different remarks on the subject. In particular, it explains Moore's well-known isolation test for intrinsic value from *Principia Ethica* (sect. 112). It also explains why Moore takes being *exactly alike* as equivalent to having the same intrinsic nature (i.,e., having the same internal properties; cf. the references given in n. 5 above). Surely, whether two objects are exactly alike might in part depend on their contingent features.[7] According to our interpretation, then, intrinsic value may well supervene on the object's contingent properties, as long as these properties are internal to the object and thus are independent of its external context.[8]

1. NON-INTRINSIC FINAL VALUES – FIRST TRY

In her argument for (C), Korsgaard assumes that an object is intrinsically valuable if the "source" of its goodness lies in the object itself. Then she suggests that for things valuable as ends, the source of their value may well be external – instead of being located in the objects themselves, it may lie in our interests or desires that are directed towards the objects in question (Korsgaard 1996a (1983), p. 252 [* p. 80 of this volume]).[9] Now, it seems to us that this argument for (C) is misdirected. The

---

[7] Moore may have been misled by his choice of examples. When explicating the notion of an intrinsic nature, he tends to discuss purely perceptual objects, such as patches of colour or perceptual patterns. (Cf. Moore, 1922, pp. 261 ff. and 1993, p. 24.) For objects of this kind it is reasonable to assume that all their qualitative non-relational properties are essential to them. Needless to say, ordinary objects are different in that respect.

[8] This "context-independence" of intrinsic value has been questioned by ethical particularists. Thus, Jonathan Dancy distinguishes "between the features that form the resultance base for value [the set of properties 'in virtue' of which the object is valuable] (which in this case are intrinsic features of the value-bearer...) and other features whose presence or absence may affect the ability of the features in the resultance base to generate the value that they do." Since "[t]he latter features...may be features of other objects", intrinsic value may change even though intrinsic features remain constant. "Intrinsic value, then, though its resultance base may not be affected by changes in other objects, may yet itself be affected by such changes, where those changes affect the ability of the intrinsic resultance base to play [its] (perhaps normal) role." (Dancy, 1999) To put it differently, while Dancy takes intrinsic value to "result" from the internal properties of the object, he allows for the possibility that the class of properties on which intrinsic value supervenes includes both internal and relational properties. Resultance and supervenience are according to him different dependence relations (cf. Dancy, 1993, ch. 5, sections 1 and 2). In what follows, we disregard this interesting complication.

[9] Thus, Korsgaard writes: "Separating the two distinctions in goodness [intrinsic/extrinsic and final/instrumental], however, opens up another possibility: that of something which is extrinsically good yet valued as an end. An example of this would be something that was good as an end because of the interest that someone took in it, or the desire that someone had for it, for its own sake" (1996a (1983), p. 252 [* p. 80 of this volume]). As noted by Brülde (1998, p. 390), in her discussion of final value, Korsgaard does not clearly distinguish between being valued and being valuable (as an end, or for its own sake). In what follows, however, we focus on the latter.

idea of a source of value is ambiguous between at least two interpretations: on the one hand, one may be thinking of the features of an object on which its final value *supervenes* (its "good-making" properties). If these features are internal to the object, i.e., non-relational, then the relevant value is intrinsic, as we are using this term. On the other hand, one may have in mind the *constitutive grounds* of an object's final value. The latter may well lie outside the object itself even though the former are internal to the object. To give an analogy: that a certain move in chess is admissible is a feature that supervenes on the internal properties of the move and of the situation on the chessboard. But the constitutive grounds of its being admissible are to be found in something external – in our conventions that determine the game of chess. Similarly, according to some preferentialist conceptions of value, preferences or desires may bestow a value on the object towards which they are directed. Still, if the object is being preferred for the features that are *internal* to it, then this externally constituted value is intrinsic: it supervenes on the internal properties of the object, precisely those properties for which the object is being preferred (cf. O'Neill, 1992: Rabinowicz and Österberg, 1996; for a critical discussion of this form of preferentialism, see Bykvist, 1998).

If our preferential attitudes are seen as the constitutive grounds of an object's final value, then it is natural to suppose that they confer value on the object not only in the actual world but also in other possible worlds towards which these attitudes may be directed. Consequently, they might confer value on objects even in those worlds in which our attitudes towards these objects would have been different from what they actually are.[10] Thus, unlike the supervenience base, the constitutive grounds need not be present in all those possible worlds in which the object is supposed to be valuable (cf. Rabinowicz and Österberg, 1996, pp. 10 and 22 f.).[11]

Claims about the supervenience bases of value belong to axiology, while claims concerning the constitutive grounds are perhaps best seen as belonging to metaethics, even though the boundary between these two disciplines is not as clear-cut as one might wish.[12] One might also try to express the distinction between supervenience and constitution in yet another way. We appeal to the properties on which the final value of the object supervenes as *reasons* for or against various

[10] In this respect, the analogy with chess is not perfect. A move that is permitted given our chess conventions would be forbidden if the conventions were different. While preferential attitudes as grounds for value are supposed to have an unlimited modal scope, chess conventions do not apply to all possible worlds.

[11] Michael Zimmerman (in personal communication) has objected to this "modal" way of characterizing the difference between the supervenience base and the constitutive grounds of value. As he points out, if our preferential attitudes in the actual world w do confer value on an object a in various possible worlds – not just in w, but also in those possible worlds in which a is not being preferred – then it still remains true that a in all these worlds has a "world-indexed" property of being-preferred-in-w. Thus, *if* one admits the existence of such world-indexed properties, one might still insist that a's value in a possible world supervenes on its property of being preferred in the actual world. For the reasons given below, however, we do not find this view especially attractive.

[12] For a short discussion of this suggestion, due to Torbjörn Tännsjö, and for some possible objections to it, see Rabinowicz and Österberg (1996), pp. 25 f.

actions taken for the sake of the object. But our preferences or desires that are directed towards the object cannot be appealed to as such reasons, at least not as long as it is a question of acting for the object's own sake. They must remain in the background rather than in the foreground of our motivational set.[13] They can be the constitutive grounds of final value but this does not make them a part of its supervenience base.

It seems, then, that Korsgaard's claim that values as ends need not be "intrinsic" boils down to the assertion that the constitutive grounds of final values may well be external. This is still perfectly compatible with the view that all values as ends are intrinsic in our sense of the term, i.e., that they always supervene on the internal properties of their bearers. This apparent conflation of the supervenience base with the constitutive grounds still persists in Korsgaard (1998), where she reiterates the claim that our particular ends "have only extrinsic value, since their value depends on our own desires and interests in them and is conferred on them by our own rational choices" (1998, p. 63).

Brülde (1998, pp. 388 f.) also suggests that Korsgaard's idea of a source of value is ambiguous between two readings, but the ambiguity he identifies is different from the one we focus on. To say that an object has the source of its value in itself may mean, according to Brülde, either that the value supervenes on the object's intrinsic features *or* that its value is *non-derivative*, i.e., that it does not "inherit" its value from some other valuable object. It seems clear that this conflation is also present in Korsgaard's 1983 paper, along with the one pointed out above. According to Korsgaard's neo-Kantian view, the final value of our ends is derived from the intrinsic value of the good will, or "humanity" (= rational choice), that has these ends. "[G]oodness, as it were, flows into the world from the good will, and there would be none without it." (1996a (1983), p. 259 [* p. 86 of this volume]). It should be noted that Korsgaard in her recent writings has given up her original conviction that what confers value (= the constitutive ground of final value) must *ipso facto* be valuable itself. She no longer believes that conferring value is some form of value transfer. Cf. Korsgaard (1996b) and (1996a, p. 407). As she puts this in her reply to Jerome Schneewind's criticisms: "More recently, especially in *The Sources of Normativity*, I have come to think of the value we place on ourselves as also conferred. To this extent, I agree with Schneewind that there is a continuity between the value of humanity and the value of other things: they are all the result of our own acts of conferring value... I am aware that in early papers I made it sound too much

[13] See Pettit and Smith (1990) for a useful distinction between two potential roles of desire in decision making. Desires always belong to the motivating background of an agent's decision but only seldom figure in the foreground of his deliberation, as reasons that the agent would appeal to. It should be noted, however, that Korsgaard herself tends to assign a foregrounding role to our conative attitudes towards final values. In her discussion of "the instrumental principle", which requires us to take the means to our ends, she argues: "for the instrumental principle to provide you with a reason [to take the means], you must think that the fact that you will an end *is a reason* for the end… It means that your willing the end gives it a normative status for you..." (Korsgaard, 1997, pp. 245-6). This view of Korsgaard, that willing an end can be one's reason for pursuing it, is convincingly criticised in Broome (1999).

as if value were some sort of metaphysical substance that gets transferred from us to our ends via the act of choice." (1998, pp. 63 f.) [14] But, as previously mentioned, she still believes that the value of a thing must be extrinsic if it is conferred on it from the outside, by our rational attitudes and choices.

2. NON-INTRINSIC FINAL VALUES – SECOND TRY

The conflation of the constitutive grounds of value with its supervenience base is also to be found in Kagan's paper. If intrinsic value is taken to be based solely on the object's internal properties, then, according to Kagan, value subjectivism would imply that intrinsic value does not exist. Nothing can be intrinsically valuable on "a radically subjectivist conception of value, according to which nothing would be valuable as an end in the absence of there being some creature who values it" (Kagan 1992, p. 184). Just as Korsgaard, Kagan here conflates the supervenience base of value with its constitutive ground: it is the latter, not the former, that according to the subjectivist conception is located in the creatures that value the object.

However, Kagan also provides another argument for (C) that is more relevant to our present concerns. The same argument is put forward by O'Neill. Both of them suggest that a thing may have value for its own sake in virtue of its *relational* features, such as uniqueness (Kagan, 1992, p. 184) or rarity (O'Neill, 1992, p. 124). A rare or unique thing may be valued even when this thing is neither a means to something valuable nor a part of some larger valuable whole. "What is valued for its own sake might well be valuable in part because of various relational properties." (Kagan, ibid.) [15] Note that, in this case, the final value of an object is supposed to *supervene* on a relational property: we value a thing *on account of* its being rare or unique. Thus, it is no longer a question of conflating the constitutive grounds of value with its supervenience base. Whether such a valuation is plausible or not is a

---

[14] In what follows, the notion of "non-derivative" value will not be further discussed. Before we leave this issue, let us just mention that Brülde (ibid., pp. 7 and 392) takes non-derivative values to be a proper subset of final values. What is non-derivatively valuable is valuable for its own sake, but the opposite does not hold. For example, a whole may be valuable for its own sake and still it may be merely derivatively valuable, if its value derives from the value of its parts.

[15] For an early statement of a somewhat similar view, see Beardsley, 1965, pp. 1 f. [* pp. 61 ff. of this volume]. According to Beardsley, an object such as a rare stamp is valuable on account of its relational property (rarity), without for this reason being valuable for the sake of something else. Unlike Kagan and O'Neill, however, he does not want to draw the conclusion that an object of this kind is valuable for its own sake. Why he hesitates to make this step is unclear to us. It may well be the case that the distinction between value for its own sake and for the sake of something else may not be exhaustive. For all we know, there might be values that are neither. But why should the value of rare stamps belong to this special category? Beardsley's only argument seems to be that the stamp cannot be valued for its own sake because its value "can be taken away, without altering the stamp at all, simply by having the Post Office Department print a few hundred million more copies" (ibid., p. 1 [* pp. 61-62 of this volume]) But this simply presupposes that the valuation of an object for its own sake, if justified, must be based on the properties the object has in itself (the properties it would keep when its copies were made). However, it is precisely this presupposition that is being questioned by a proponent of claim (C).

substantive issue that need not detain us here.[16] But supposing it is plausible, what is the status of the value involved? This is the question we want to discuss. Thus, what we are interested in is a particular issue in formal theory of value.

If O'Neill and Kagan are right, it is easy to find many other cases of non-intrinsic final values. O'Neill gives an example from environmental ethics: a wilderness may be valuable in this way because of its being untouched by humans, which may be seen as a relational property (on the assumption that the complements of relational properties themselves are relational, in an extended sense). Another class of examples involves cases in which a thing is valued for its own sake in virtue of its special relationship to a particular object, event, or person. An original, say, an original work of art, may be valued for its own sake precisely because it has the relational property of being an original rather than a copy. Its final value supervenes, in part, on its special causal relation to the artist. Princess Diana's dress may be another case in point. The dress is valuable just because it has belonged to Diana. This is what we value it for.

But, one might object, is it really a case of a non-instrumental value? Diana's dress is perhaps valuable merely as a means: merely because it allows us to establish an indirect connection to a person we admire or find important in one way or another. Having such a connection may be something that we set a final value on. Couldn't this be what is going on here? Not necessarily. Even if the desire to establish such an "affiliation" with Diana may well be a part of the causal explanation of our evaluative attitude towards the dress, this does not imply that the evaluative attitude itself is of the instrumental kind: if we idolise Diana, we do not simply find the dress useful for some purpose; we ascribe an independent value to it. Compare this with O'Neill's example: the wilderness is not simply instrumental in allowing us to come into contact with something (otherwise) untouched by humans. Even if we could never visit the wild area, it would still keep its value from our point of view.

But if Diana's dress is a good example of what we are after, then there are innumerable examples of a similar kind: Napoleon's hat, a gun that was used at Verdun, etc. In all these cases, a thing acquires a non-instrumental value in virtue of its causal relation to some person, object or event that stands out in some way. The final value of a thing comes from the importance of another thing. Note that there might be a connection between this kind of final value and the final value based on rarity. If many of Napoleon's hats are still around, the value of each of them may be considerably lower.

---

[16] Michael Zimmerman (personal communication) is one of the sceptics. Why should an object such as a rare stamp be valuable for its own sake just on account of its being rare? After all, the stamp also has other relational properties which make it valuable without, however, making it valuable for its own sake. For example, the property of facilitating communication. Why should rarity have such a special status then? This is a fair worry, but we shall not dispel it here. Much the same kind of worry arises when someone asks why some of the intrinsic features of the object make it valuable for its own sake, but others don't. Questions of this kind can only be answered by a substantive axiological theory and not by a formal theory of value.

Korsgaard discusses a different type of case in which the final value of a thing depends on its relational properties: we can say "of certain kinds of things, such as luxurious instruments, that they are valued for their own sakes under the condition of their usefulness [a relational property]. Mink coats and handsome china and gorgeously enamelled frying pans are all things that human beings might choose partly for their own sakes under the condition of their instrumentality: that is, given the role such things play in our lives" (Korsgaard, 1996a (1983), p. 264 [* p. 89 of this volume]). It is an interesting and somewhat paradoxical group of cases: the final value of a thing is here supposed to supervene, in part, on its instrumental value (or, alternatively, on the same relational property that makes the object instrumentally valuable). Being useful is, of course, not sufficient to make the mink coat valuable for its own sake; but it is a necessary pre-condition of the coat's final value.

3. REDUCTION MANOEUVRE

The ascription of value to an object on the basis of its relational properties can be met with the following response: when you say that an object, X, is valuable for its own sake because of its relation to some other object or objects, what you really mean is that the value accrues to a *state* (or a fact) to the effect that the object X that has this relation exists. This valuable state contains the relevant relation as its component and so its value may be seen as *intrinsic*: the value of the state is based on the internal features of the state itself. For example, the value of the wilderness untouched by human hands is, according to this proposal, nothing but the intrinsic value of the state *The wilderness untouched by human hands exists.* Analogously, to say that Diana's dress is valuable for its own sake reduces to the claim that the state *Diana's dress exists* is intrinsically valuable. What was the external, relational property of the dress (its having belonged to Diana) is now seen as the internal feature of the state. Consequently, the value of the state is based on its internal properties and therefore is intrinsic (in the standard Moorean sense).

If this reduction is correct, there is no reason to assume the existence of final values that are not intrinsic. The statement that something has a final value on the basis of its relational properties is always reducible to the claim that some state that involves the object in question together with its relations to other objects is intrinsically valuable.

But is such a reduction reasonable? At first sight, it does not seem to be: it appears to put the cart before the horse. If the existence of the wilderness is valuable, it is *because* the wilderness itself is valuable. Thus, it is the state that derives its value from the object it involves and not the other way round.[17] Consequently, the value of the object does not seem to be explicable in terms of the value of the state.

[17] Anderson (1993, p. 26) goes as far as to suggest that, *in general*, "states of affairs which consist in the existence of something are valuable only if it makes sense to care about the thing that exists". To make the point we are making we need not subscribe to this general view.

In fact, in some cases, one might even want to go further and question whether the existence of a valuable thing is always something to be valued. Does it follow that the existence of X is a valuable state, if X is a valuable object? As is well-known, this entailment was emphatically denied by Kant in connection with aesthetic value: the judgment that the object is aesthetically valuable does not in the least commit one to the conclusion that the object's existence is of value (Kant, 1913 (1790), § 2). It may well be the same with some judgements to the effect that an object is valuable for its own sake. We take Diana's dress to be valuable, but do we think it is valuable that this dress exists? We might hold that, because of its value, the dress should be preserved or that it should not be destroyed, but do we think that the world is a better place because of the existence of that object? [18]

This shows, by the way, that the friends of the reduction might have a difficult time in identifying the precise state which is supposed to be valued. Is it the existence of the thing in question or is it rather its continued existence given that it already exists? Or perhaps it is not the continued existence of a thing that is good, but rather its ceasing to exist that would be *bad*? Or perhaps it would be bad only if the thing in question were wilfully destroyed? There are here several possibilities to choose between and it is not at all clear that there is just one right candidate.[19]

We may at this point, however, take a pause and ask a more fundamental question: why should anyone, in the first place, want to reduce the value of a thing to the value of a state? Several philosophers (Ross, 1930; Harman, 1967; von Wright, 1963 and 1972; Chisholm, 1968-69; Quinn, 1974; Oldfield, 1977; Quinn, 1977; Carlson, 1995 and 1997; Feldman, 1997; Bradley, 1998; Brülde, 1998; and Zimmerman, 2001) argue for or simply assume that nothing but states (or facts) can be ultimately valuable. But what can be the motivation for such a radical position?

One reason (or at least one explanation) has to do with familiarity. The logic and the semantics of such modal propositional operators as "It is good that..." or "that __ is better than that ..." has undergone an impressive development during the last forty years of this century. A value theorist who reduces all value to proposition-like objects can therefore fall back on a rather well-developed formal framework. Another reason has to do with simplicity. It is widely accepted that some states *are*

---

[18] See also Anderson (ibid.): "One may suppose that it doesn't make sense to care about anything unless it makes sense to care about its existence... But the supposition is not true. It may make sense for me to love a person, but this does not imply that I must want that person to continue living. If he is gravely ill, it may be the best expression of my love for him to wish that he die quickly and mercifully."

[19] Another reduction possibility has been suggested to us by Ingmar Persson. That an object X is valuable in virtue of its property P amounts according to him to the claim that it is valuable *that X has P*. At first sight, this seems like the old proposal in a new disguise: *that X has P* is logically equivalent to the state *that X, which has P, exists*. Still, as Persson points out, "the cart before the horse"-type of objection does not apply to his proposal: while the existence of X, which has P, may be valuable simply because X, which has P, is a valuable object, the same kind of response cannot be used with respect to the state that consists in X having P. However, even if we accept this point and thus reject logical equivalence as the sufficient identity criterion for states, we might well wonder whether the proposed reduction is intuitively plausible. We take the dress to be valuable because it has belonged to Diana, but do we think it to be a valuable state of affairs *that* the dress has belonged to her? Why should one subscribe to such a strange value claim?

intrinsically valuable. Consequently, a view that reduces all final value to the intrinsic value of states has the obvious advantages of monism. In particular, computing the values of the wholes on the basis of the values of their parts is greatly facilitated if all their valuable parts have the same ontological status. Aggregating the values of things of different sorts with the values of states might prove to be an overwhelming task. Also, if we reserve all value to states, we might hope to trace all such value to the value of some "basic states" that only contain "value-making" components. A similar removal of irrelevant features is, however, impossible as far as the things rather than the states are concerned: obviously, a thing cannot be stripped of a large number of its features and still exist.

But simplicity is a two-edged sword in the hands of a theorist: the theory itself is easier to handle but it may have difficulties in capturing the complexities of its subject-matter. A more compelling reason for the reduction might have to do with some influential views concerning the nature of value. According to one such group of views, there exists an essential connection between value and preference, or desire. Values are seen as accruing to objects of preferences. To be valuable is to be desirable, to be *a fitting object of a preference* (or, in a more subjectivist version of such a preferentialist position, it is to be an object of a preference we would have under certain ideal conditions). Now, it seems plausible to say that preferences have states as their objects and not concrete things. In those cases when we talk about preferring a thing, for example preferring an apple to an orange, it is really the having or the eating of an apple that is being preferred.[20] But if objects of preferences are states and being valuable is being a fitting object for a preference, then only states can be valuable.[21] A similar conclusion is reached if value is analysed in terms of pursuing or promoting. We promote states. Consequently, if value is *what we should promote* (or what we would promote under ideal circumstances), then – again – it is only states that are valuable. As Ross puts it, "'good' is the name of a quality which attaches ... only to 'objectives', and not to 'substances'". Since "objectives" always are facts, he concludes that it is only facts that can have ultimate value (Ross, 1930, pp. 112 f.).

However, even if we accept the general idea that value is what calls for an appropriate response, the response in question need not consist in just preferring or

[20] Instead of letting states be the objects of preference, we might interpret (some or all) preferences as *attitudes de se* that take one's own properties as objects (properties such as having an apple or eating an apple). Cf. Lewis (1979, esp. section 9). The difference between the "state view" and the "property view" is not really important in the present context; both views deny that preference can take concrete individuals as objects.

[21] But see Hansson (1998, p. 63): "It is not quite as simple as that. Some preferences are difficult to reconstruct with states of affairs as relata. As was pointed out to me by Wlodek Rabinowicz, particularly good examples of this can be found in the aesthetic realm. The...statement that Bartok's fourth quartet is better than his third cannot be satisfactorily expressed as a preference for the state of affairs that the fourth quartet exists (is played, is listened to) rather than the third". It is implicitly assumed in this counter-example that "better than" should be understood as "is preferable to". Consequently, the effectiveness of the counter-example crucially depends on whether value coincides with preferability. For a critical discussion of this assumption, see below.

promoting. There may well be other alternatives: preference is not the only attitude to be considered, nor is promoting the only behaviour that may be relevant in this context. Historically, there is the Brentanian tradition according to which values are interpreted as fitting objects of *emotion*: love, liking, admiration or respect. (See e.g., Brentano, 1969 (1889)). Alternatively, one might relate value to thing-oriented behaviours, or perhaps better, to thing-oriented attitudes-cum-behaviours: value is what we should cherish, protect or care for. (Note, by the way, that some analyses of value in terms of appropriate emotional responses may also involve a behavioural component: there is no clear cut between feeling and behaving.) In fact, when one realises how many various types of responses could be relevant in this context, it becomes tempting to draw the conclusion that any monistic analysis of value in terms of one particular type of response would be inadequate. Marcia Baron has instead proposed a *pluralist* treatment: a particular value may call for several different types of response, and – what is especially relevant to our concerns – different values may call for different responses (Baron 1997; for closely related pluralist views, cf. Anderson, 1993, ch. 1, section 1, and Swanton, 1995.[22]). As Baron puts it:

> Value comes in many varieties, even if we limit ourselves...to non-instrumental value, and it doesn't appear that all value calls for the same response. Some are such that the best response is to exemplify or instantiate them; still others call for producing as much of them as possible; others call for honoring them by refraining from doing anything that would violate them. A mixture of these responses will often be called for, a mixture whose proper proportions may differ, depending on the value and the particular situation. (Not that there will typically be only one appropriate response or blend of responses; there will be some inappropriate ones, but often more than one appropriate response.) (Baron, 1997, p. 22)

Given such a pluralist approach to value analysis, the main motivation to reduce thing values to state values disappears. Valuable things may be objects that call for specific thing-oriented attitudes or behaviors: a wilderness untouched by human hands calls for protection, Diana's dress is an object to be cherished and preserved, and so on.

In fact, there are theorists who connect *all* value with non-propositional, thing- or person-oriented attitudes, such as love, respect, honouring, etc. Elisabeth Anderson is a case in point (Anderson, 1993). Anderson therefore draws the radical conclusion that ultimate value only accrues to non-propositional entities (ibid., pp. 20 ff). We need not go that far: our preferred version of the pluralist approach makes room both for propositional and non-propositional value responses.

Admittedly, if all the different attitudinal and behavioural responses to value could in *their* turn be reduced to preference, promotion or some combination of the two, the reduction of all value to the value of states would again become viable. But

[22] Baron is strongly influenced by Christine Swanton's paper. Unlike Baron, however, Swanton does not allow that different values might require different response configurations (cf. Swanton, 1995, p. 49).

this avenue seems quite unpromising: the relevant attitudes and behavior seem to be too complex to allow such a reconstruction.[23]

Even if non-propositional objects are admitted as bearers of intrinsic value, one might still insist that value as an *end* can only accrue to proposition-like entities, such as states of affairs or facts.[24] After all, ends are never things. Nor are they ever persons, *pace* Kant. They are *objectives* and Ross was right in his insistence that objectives are facts rather than "substances" (Ross, ibid.).[25] This means, however, that the term "value as an end" may not be quite suitable for the kind of value we have in mind when we say that a non-propositional object, such as Diana's dress or a wilderness untouched by human hands, is valued for its own sake. While a "sake"[26] may be an end, a non-propositional object can't be, even when it is valued for its own sake. Still, whereas "value as an end" is for this reason misleading, the relevant values can be said to be "end-point values", insofar as they are not simply conducive to or necessary for something else that is of value. They are "final", then, in this sense of being "ultimate".

There is a further issue lurking in the background; once we distinguish between the intrinsic and the final, the standard *ad hominem* regress argument for the existence of intrinsic values turns out to be based on an equivocation. The argument goes like this: Clearly, there are things we take to be valuable as means. But nothing would be valuable as means, i.e., for the sake of something else, unless there were things that are valuable for their own sakes. The sequence: X is valuable because it is a means to some Y that is valuable as a means to some Z that is valuable as a means to... must, it seems, sooner or later come to an end, if we are to avoid a vicious regress. Now, however, it should be clear that this argument, if at all correct, only proves that we have to admit the existence of *final* values. That *intrinsic* values exist cannot be shown by any comparable regress argument that starts from the existence of extrinsic values.[27]

---

[23] See Velleman, 1999, pp. 353 f., for an attempt to show that such an attitude as love cannot be given a conative analysis in terms of some states of affairs that a loving person by necessity is supposed to pursue.

[24] As we use the term, a "proposition-like" entity may or may not be abstract; it may or may not be spatiotemporally located and its identity criteria may be more or less stringent. The important consideration in the present context is that such entities can be objects of so-called "propositional attitudes" and thus can be referred to by means of "that"-clauses.

[25] But cf. Velleman (1999), pp. 355 f., where it is suggested that the philosophers who interpret "ends" as synonymous with "aims" or "objectives" are mistaken. The end of an action is declared by Velleman to be "that for the sake of which one acts", which clearly need not be the same as the aim of that action (if it has an aim). This rather doubtful move is used by him to defend Kant's view of persons as ends. But are we really prepared to say that the person for the sake of whom we act is the *end* of our action?

[26] The word "sake" has the same origin as the German "Sache" or the Swedish "sak". Thus, to act for someone's sake is, literally, to act for his "Sache", which may well be an end of some sort.

[27] *Pace* Korsgaard (1996a (1983), p. 259 [* p. 86 of this volume]). Korsgaard, who identifies extrinsic value with Kant's notion of a conditional good, mistakenly suggests that the regress on the conditions of conditional goodness must sooner or later lead us to something that is unconditionally good. Clearly, this is a non-sequitur: it is not logically necessary for the condition of goodness to be good itself, either conditionally or unconditionally.

As we pointed out at the beginning of this paper, the intrinsic value of an object may well vary in different possible worlds, if this value supervenes on the object's *contingent* internal properties. In other words, one should not confuse intrinsic value with *essential value*, where the latter is the value that the object has by necessity, i.e., in every possible world in which it exists. This suggests, by the way, that the notion of an intrinsic value may not be as normatively interesting as many have thought. After all, what is so special about value that supervenes on the object's internal rather than relational properties, especially if the former may be just as contingent to the object as the latter?[28] One can easily see the normative relevance of the notion of a final value (at least if such value is analysed in terms of a range of the fitting responses that the object calls for), but the concept of an intrinsic value seems to lack a special normative interest. (See Brülde, 1998, pp. 389 f.) This observation suggests that the identification of final value with intrinsic value is perhaps not even *prima facie* very convincing. Its seeming immediate appeal may depend on our carelessness with language. The inference from "valuable for its *own* sake" to "valuable (as it is) *in* itself" is quick but invalid.

In conclusion, it seems to us that the case for irreducible final but non-intrinsic values is quite strong. Final values that accrue to things in virtue of relational features cannot be traced back to the intrinsic values of states that involve these things together with their relations. On the contrary, such states, insofar as they are valuable at all, seem to derive their value from the things involved. The endeavour to reduce thing-values to state-values is largely motivated by a mistaken belief that the appropriate responses to a valuable object at bottom always consist in preferring and/or promoting. A pluralist approach to value analysis obviates the need for reduction: the final value of a thing can be given an independent interpretation in terms of the appropriate non-propositional responses the thing in question calls for.[29]

---

[28] In fact, this may explain Moore's tendency to ascribe necessity to internal properties. See our discussion above, at the end of the introductory section.

[29] As we have recently discovered, the views put forward in this paper show several similarities to the position defended in Kagan (1998) [* pp. 97-114 of this volume]. Admittedly, Kagan still continues to conflate the constitutive grounds of value with its supervenience base: he insists that, if subjectivism were true, all value would partially supervene on relational properties. Otherwise, however, his examples of final values are very similar to ours. He especially elaborates on a point that we only make in passing, in connection with Korsgaard's example of a mink coat: in some cases, the final value of an object can supervene on its value as means. Kagan also notes that intrinsic value may well be contingent, and he draws the same conclusion as we do: it is the final values that are interesting from the normative point of view. In fact, he even takes this as a reason for a linguistic reform: he suggests we should from now on use "intrinsic value" to denote final value. We do not want to go that far. Finally, Kagan also takes up the reduction manoeuvre, i.e., the suggestion that values of things are reducible to values of states. As he notes, this move would allow one to insist that, in the end, all value for its own sake must supervene on internal properties.

Kagan, however, does not adduce any reasons to resist the reduction move. In fact, he seems to believe that such a reduction might well be successful. In this decisive respect, Kagan's position differs from ours. In our view, there would not be much point in the argument for final non-intrinsic values if such values after all could be reduced away to the intrinsic values of states.

REFERENCES

Anderson E., 1993. *Value in Ethics and Economics*. Cambridge, Mass.: Harvard University Press.

Baron Marcia W., 1997. “Kantian Ethics”, in *Three Methods of Ethics*, Marcia W. Baron, Philip Pettit, and Michael Slote. Oxford: Blackwell.

Beardsley, Monroe C., 1965. “Intrinsic Value”. *Philosophy and Phenomenological Research*, 26: 1-17 [* pp. 61-75 of this volume].

Bradley, Ben, 1998. “Extrinsic Value”. *Philosophical Studies*, 91: 109-26.

Brentano F., 1969 (1889). *The Origin of Our Knowledge of Right and Wrong*. London and Henley: Routledge & Kegan Paul.

Broome, John, 1999. “Normative Requirements”. *Ratio*, 12: 398-419.

Bykvist, Krister, 1998. *Changing Preferences: A Study in Preferentialism.* Uppsala: Acta Universitatis Uppsaliensis.

Brülde, Bengt, 1998. *The Human Good*. Göteborg: Acta Universitatis Gothoburgensis.

Chisholm, Roderick M., 1968-9. “The Defeat of Good and Evil”. *Proceedings and Addresses of the American Philosophical Association*, 42: 21-38.

Carlson, Erik, 1995. *Consequentialism Reconsidered.* Dordrecht: Kluwer.

Carlson, Erik, 1997. “The Intrinsic Value of Non-Basic States of Affairs”. *Philosophical Studies*, 85: 95-107 [* pp. 361-70 of this volume].

Dancy, Jonathan, 1993. *Moral Reasons.* Oxford: Blackwell.

Dancy, Jonathan, 1999. “On the Logical and Moral Adequacy of Particularism”, *Theoria*, 65: 212-24.

Feldman, Fred, 1997. “On the Intrinsic Value of Pleasures”. *Ethics*, 107: 448-466.

Hansson, Sven Ove, 1998. *Structures of Value – An Investigation of the Statics and Dynamics of Values and Norms.* Lund Philosophy Reports, Department of Philosophy, Lund University.

Harman, Gilbert H., 1967. “Toward a Theory of Intrinsic Value”. *Journal of Philosophy*, 64: 792-805 [* pp. 349-60 of this volume].

Humberstone, I. L., 1996. “Intrinsic/Extrinsic”. *Synthese*, 108: 205-67.

Kagan, S., 1992. “The Limits of Well-being”, in *The Good Life and the Human Good,* eds. Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul. Cambridge: Cambridge University Press.

Kagan, S., 1998. “Rethinking Intrinsic Value”. *Journal of Ethics*, 2: 277–97 [* pp. 97-114 of this volume].

Kant, I., 1913 (1790). *Kritik der Urteilskraft*. Akademieausgabe, Berlin: Verlag Georg Reimer.

Korsgaard, Christine M., 1996a. “Two Distinctions in Goodness”, in Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press): 249-74; reprinted from *Philosophical Review*, 92 (1983): 169-95 [* pp. 77-96 of this volume].

Korsgaard, Christine M., 1996b. *The Sources of Normativity.* Cambridge: Cambridge University Press.

Korsgaard, Christine M., 1997. “The Normativity of Instrumental Reason”, in Garrett Cullity and Berys Kraut (eds), *Ethics and Practical Reason* (Oxford: Oxford University Press): 215-54.

Korsgaard, Christine M., 1998. “Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer and Schneewind”. *Ethics*, 109: 49-66.

Lemos, Noah M., 1994. *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge University Press.

Lewis, David, 1979. “Attitudes *De Dicto* and *De Se*”. *Philosophical Review*, 88: 513-43.
Moore, G. E., 1903. *Principia Ethica*. Cambridge: Cambridge University Press (revised edition, 1993).
Moore, G. E., 1922. *Philosophical Studies*. London: Routledge and Kegan Paul.
Moore, G. E., 1965 (1912). *Ethics*. Oxford: Oxford University Press.
Moore, G. E., 1993. “Preface to the second edition”, in *Principia Ethica*, revised edition. Cambridge: Cambridge University Press.
Oldfield, Edward, 1977. “An Approach to a Theory of Intrinsic Value”. *Philosophical Studies*, 32: 233-49.
O’Neill, John, 1992. “The Varieties of Intrinsic Value”. *The Monist*, 75: 2: 119-137.
Pettit, Philip and Smith, Michael. 1990. “Backgrounding Desire”. *Philosophical Review*, 99: 565-92.
Quinn, Warren, 1974. “Theories of Intrinsic Value”. *American Philosophical Quarterly*, 11: 123-132.
Quinn, Philip, 1977. “Improved Foundations for a Logic of Intrinsic Value”. *Philosophical Studies*, 32: 73-81 [* pp. 241-48 of this volume].
Rabinowicz, Wlodek and Österberg, Jan, 1996. “Value Based on Preferences”. *Philosophy and Economics*, 12: 1-27.
Ross W. D., 1930. *The Right and the Good*. Oxford: Oxford University Press.
Swanton, Christine, 1995. “Profiles of the Virtues”. *Pacific Philosophical Quarterly*, 27: 335-344.
Tolhurst, William, 1983. “On the Nature of Intrinsic Value”. *Philosophical Studies*, 43: 383-95.
Velleman, J. David, 1999. “Love as a Moral Emotion”. *Ethics*, 109: 338-75.
von Wright, G. H., 1963. *The Varieties of Goodness*. London: Routledge and Kegan Paul.
von Wright, G. H., 1972. “The Logic of Preference Reconsidered”. *Theory and Decision*, 3: 140-169.
Zimmerman, Michael J., 2001. *The Nature of Intrinsic Value*. Lanham: Rowman and Littlefield.

CHAPTER 11

J. J. THOMSON

THE RIGHT AND THE GOOD*

There is a story about the right and the good which strikes me as very attractive.[1] Although there are gaps in it, I think its structure must be correct. But first some background.

The story about the right and the good which comes down to us from G. E. Moore – I shall call it *Moore's story* – is familiar enough, and many people find it very attractive. Many people nowadays reject it, but even among them, many understand perfectly why its friends resist invitations to reject it: that is because they themselves accept its first two parts and concede that it is hard to supply a clear rationale for rejecting its concluding third part.

Moore's story begins with the good. Some things are good, Moore said, and some things are not good; so there is such a property as goodness – all good things have it and all things that are not good lack it.[2] Moore had much to say about which things do and which things do not possess the property goodness, but for a reason

---

* An early draft of this article was presented at the Chapel Hill Colloquium in 1995, and later drafts were presented as the Hägerström Lectures for 1996 at the University of Uppsala, and at a symposium at University College, London; I am grateful to the participants for their comments. The later drafts were written at the Centre for Advanced Study in Oslo; I am grateful to the Centre for providing me with support and very pleasant surroundings in which to work. Many people gave me helpful comments along the way: I thank Alexander Byrne, James Higginbotham, Ralph Wedgwood, and, in particular, Catherine Elgin, who commented on several drafts.

[1] I shall be trying to improve on the story I told in chapter 8 of Gilbert Harman's and my *Moral Relativism and Moral Objectivity* (Cambridge: Blackwell, 1996).

[2] Moore said in *Principia Ethica*: "Ethics is undoubtedly concerned with the question what good conduct is; but, being concerned with this, it obviously does not start at the beginning, unless it is prepared to tell us what is good as well as what is conduct. For 'good conduct' is a complex notion: all conduct is not good; for some is certainly bad and some may be indifferent And on the other hand, other things, besides conduct, may be good; and if they are so, then, 'good' denotes some property, that is common to them and conduct; and if we examine good conduct alone of all good things, then we shall be in danger of mistaking for this property, some property which is not shared by those other things..." – Thomas Baldwin, ed. (New York: Cambridge, 1993, rev. ed.), p. 54. Thus, good conduct is what has the two properties goodness and being conduct, and other things besides good conduct have the property goodness: goodness is the property that all good things have in common.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 131-152.
© 2005 *Springer. Printed in the Netherlands.*

that will emerge, I shall not summarize his views on this matter. The second part of the story flows from the first: there being such a property as goodness, there is also such a relation as being better than, or *betterness.* (Not every property has a comparative, but this one surely must.) Moore's story then concludes: the right is analyzable in terms of the relation betterness. Thus, for it to be the case that Alfred ought to do a thing at a time is for it to be the case that the world will be better if he does the thing than it will be if he does any of the other things it is open to him to do at the time.

There are at least three things that make this story seem attractive. In the first place, it is wonderfully simple. It counts in favor of a theory if it supplies a simple picture of what had initially looked like chaos. Second, it does seem to us that there is such a thing as one thing's being good and another's not being good, and that there is such a thing as the world's being better if things go this way rather than that. We certainly say *words* like this often enough. Third, and most important, it is hard to see how it could be perfectly all right to fail to do what would make the world go better than it otherwise would. If you really would make the world go worse if you did a thing, then are you not called on morally to refrain from doing it? What else is there for the right to turn on than how good the world will be if you act in this way rather than that?

As I said, many people nowadays reject Moore's story. The ground most commonly given for doing so is that accepting the story yields intuitively implausible conclusions about the right. Suppose that Alfred is under the following threat by the Mafia: kill Bert, or we shall kill Charles, David, and Edward. Moore's opponents say: under those conditions, the world will surely be better if Alfred kills Bert than if he leaves the Mafia to kill the three others. After all, deaths are surely bad, and three of them surely three times as bad as one. But, they go on: moral intuition delivers, firmly, that it would be wrong for Alfred to kill Bert. In short, what is wrong with Moore's story lies in its concluding part. All is well with its two earlier parts, which tell us about goodness and betterness; the trouble lies only in the story's then going on to declare that the right is analyzable in terms of betterness.

Moore's friends – and he still has a good many – reply: "But look, you have conceded that the world really will be better if Alfred kills Bert; so how could it possibly be wrong for him to do so? What is there that could be thought to make his doing so wrong, given that the world will be better if he does it?" This, in effect, asks "What's your alternative story?" and expresses skepticism about the possibility of any plausible alternative. There is much to be said for that skepticism, for it *is* hard to see what could make Alfred's killing Bert wrong if it is not bad, and then hard to see how the Mafia's killing the three other people could fail to be at least three times as bad.

This dispute still goes on in moral philosophy, and what is puzzling about it is that a deeper objection to Moore's story not only is available but has been available for many years now. Peter Geach (in 1956), Paul Ziff (in 1960), and G. H. von

Wright (in 1963)[3] gave the excellent advice that we should *look and see* how the word 'good' is actually used. They showed, conclusively, that it does not function in the way in which adjectives like 'visible' and 'happy' do. In saying 'That's good', we are not ascribing a property goodness – indeed, there is no such thing. Curiously enough, this idea was not picked up by moral philosophers, and brought to bear on Moore's story, for some twenty years. My impression is that Philippa Foot's article "Utilitarianism and the Virtues" (1983)[4] is the first work by a moral philosopher to declare that since there is no such property as goodness, Moore's story is incoherent from the outset.

It should really have been clear back in 1903 that there was something fishy about Moore's story. At the beginning of *Principia Ethica*, Moore says that the question he will be addressing himself to in what follows is the question 'What is good?', and he rightly thinks that we are going to need a bit of help in seeing exactly what question he is expressing in those words. He proposes to help us by drawing attention to a possible answer to the question he is expressing – that is, to something that would be an answer to it, whether or not it is *the* or *a* correct answer to it. Here is what he offers us: "Books are good." Books are good? What would you mean if you said "Books are good"? Moore, however, goes placidly on: "though [that would be] an answer obviously false; for some books are very bad indeed" *(op. cit.*, p. 55). Well, some books are bad to read or to look at, some are bad for use in teaching philosophy, some are bad for children. What sense could be made of a person who said, "No, no, I meant that some books are just plain *bad things*?" There is something weird about that passage in *Principia*, and it is puzzling that this was not noticed at the time.

In any case, it is should he clear enough by now that all goodness is goodness in a way, and that, if we do not know in what way a man means that a thing is good when he says of it "That's good", then we simply do not know what he is saying of it. Perhaps he means that it is good to eat, or that it is good for use in making cheesecake, or that it is good for Alfred. If he tells us, "No, no, I meant that it is just plain a *good thing*," then we can at best suppose he is a philosopher making a joke.

The same is true of betterness: it, too, is always betterness in a way. People do say the words "This is better than that", but what they mean is always that the first

---

[3] See Geach, "Good and Evil," *Analysis*, XVII (1956): 33-42, reprinted in *Theories of Ethics*, Philippa Foot, ed. (New York: Oxford, 1967); Ziff, *Semantic Analysis* (Ithaca: Cornell, 1960); von Wright, *The Varieties of Goodness* (London: Routledge, 1963).

[4] Foot's article was first published in *Proceedings and Addresses of the American Philosophical Association*, LVII (1983): 273-83. A revised version appeared in Mind, XCIV (1985): 196-209, and was reprinted in *Consequentialism and Its Critics*, Samuel Scheffler, ed. (New York, Oxford, 1988). I attribute the declaration to Foot's article and not to Foot herself, since while I continue to believe that the article can he interpreted as making it, Foot has recently said that she did not mean to do so, and she must be allowed final say as to her intentions.

It is worth mention that the idea at least *lurked* in von Wright. That exceedingly interesting book unfortunately had no impact on moral philosophy at the time; I attribute that fact to the seductiveness of Moore's story.

thing is better to eat, or better for use in making cheesecake, or better for Alfred, and so on.

But if there is no such property as goodness, and no such relation as betterness, then, a fortiori, the right is not analyzable in terms of the relation betterness, and we need not appeal to moral intuition to justify rejecting Moore's story. In sum, what is wrong with Moore's story is not merely its concluding part but its very beginning.

We need a new story about the right and the good, then. Here is the one that strikes me as attractive.

1.

It begins with what I shall call the *first-order ways of being good.* I mentioned being *good to* eat; there are also being good to look at, or listen to. I mentioned being *good for use in* making cheesecake; there are also being good for use in hammering in nails, or planting bulbs. I mentioned being *good for* Alfred; there are also being good for Alfred's lawn mower, or Alfred's apple tree. There are others, too: being *good at* hanging wallpaper, playing chess, or singing, being *good in Hamlet* (the play), being *good as* Hamlet (the Prince), being *good with* children, and so on. Let us follow a practice common among linguists in using the term 'adjuncts' to refer to the expressions 'to eat', 'for use in making cheesecake', and so on, which I added to 'good' to get 'good to eat', 'good for use in making cheesecake', and so on. In general, then, whenever we predicate an expression of the form 'good plus adjunct', we ascribe a first-order way of being good.

Perhaps the following will suggest itself: so after all there really is such a property as goodness! – it is the disjunction of the properties being good to eat, being good for use in making cheesecake, being good for Alfred, and so on. Or more compactly, that being good is being good in at least one of those ways. No doubt there is a property that is the disjunction of all of those properties. But it is an uninteresting property, for everything has it: everything is good in one or other of those ways. (If you find an example that strikes you as good in none of those ways, then it is sure to be good for use in a philosophical discussion of goodness.) It is therefore of no interest to any friend of Moore's story. Moreover, it is not what anybody ever means to be ascribing to a thing in saying 'That's good', so its title to be called "goodness" is at best dubious.

What about what we ascribe when we predicate an expression of the form 'good *K*'? Geach had said that 'good' is an attributive adjective, meaning by that: '*X* is a good *K*' is not equivalent to '*X* is good' and '*X* is a *K*'. Compare 'big', which is also attributive in that '*X* is a big *K*' is not equivalent to '*X* is big' and '*X* is a *K*'. (Geach said: by contrast, an adjective such as 'red' is predicative, for '*X* is a red *K*' is equivalent to '*X* is red' and '*X* is a *K*'.[5]) Indeed, he said something stronger, namely, that 'good' is "essentially an attributive adjective," meaning by that:

[5] 'Red' was not in fact well chosen for Geach's purposes, since 'red' is heavily context dependent: what we ascribe to an apple when we say 'It's red' is different from what we ascribe to the paint in a certain

> Even when "good" or "bad" stands by itself as a predicate [as, for example, in "That is good"], and thus is grammatically predicative, some substantive has to be understood; there is no such thing as being just good or bad, there is only being a good or bad so-and-so (*op. cit*., p. 65).

Thus, 'good' is like 'big' not merely in being an attributive adjective in the sense Geach indicated, but also in the following respect: just as there is no such thing as being just big, but rather only being a big so-and-so, so also there is no such thing as being just good, but rather only being a good so-and-so – in both cases, "some substantive has to be understood."

If this stronger claim were true, then we could (again) say that there is such a property as goodness: it is a relative property like bigness. (As being big is being a big *K* so also being good is being a good *K*.)

But the stronger claim is not true. The weaker claim is true: 'good' certainly is an attributive adjective in the sense Geach indicated, for '*X* is a good *K*' is not equivalent to '*X* is good' and '*X* is a *K*'. Moreover, it is certainly right to say there is no such thing as being just good. But it is not right to say that "some substantive has to be understood." If a man says "That's good," he may mean the likes of "That's a good book" or "That's a good banana"; but he may instead mean the likes of "That's good for use in making cheesecake" or "That's good for Alfred," and here no substantive needs to be understood. (In fact, for what *K* could it at all plausibly be thought that being good for use in making cheesecake is being a good *K*? For what *K* could it at all plausibly be thought that being good for Alfred is being a good *K*?)

I suggest, in addition, that our understanding of assertions of sentences of the form '*X* is a good *K*' – like our understanding of assertions of 'That's good' – rests on our knowing, from the context of utterance, in which of what I called the first-order ways of being good the speaker means that the thing is good.[6] I should think that most assertions of 'That's a good book' are made in contexts in which what is in question is which book to read, and the speaker is therefore likely to mean that the thing is a book that is good to read. But if what we had asked for was advice about which book to use in teaching elementary logic, the speaker is likely to mean, instead, that the thing is a book that is good for use in teaching elementary logic. In still other contexts, the speaker is likely to mean that the thing is a book that is good for children to read, or that it is a book that is good to look at. Similarly for 'That's a good banana': in most contexts, the speaker is likely to mean that the thing is a banana that would be good to eat, but where what we wanted was a banana for use in the display of fruit on the dining room table, he is likely to mean that it is a banana that is good to look at.

---

can when we say 'It's red'. Better choices would have been 'visible' or 'happy', which I mentioned above, or any of the following: 'sauteed' ('pureed', 'boiled'), or 'poisonous' ('nauseating'), or 'fermented' ('carbonated'), or 'acidic' ('alkaline'), and so on.

[6] Or in which of what I shall in the following section call *second-order ways of being good* the speaker means that the thing is good, for that, too, is a possibility.

The context has to tell us what a speaker means by 'That's a good *K*' (as what a speaker means by 'That's good') or we simply do not know, for the expression 'good *K*' leaves this open. The point emerges most starkly when we look at odd expressions of the form 'good *K*'. Suppose someone calls a thing a good corpse.[7] What on earth does the speaker mean? I have no idea, because I have no idea what he means to be saying the corpse is good for use in, or good to, or good for, and so on. ('Good state of affairs' and 'good event' are unlike 'good corpse': it is not that we have no idea what a speaker might mean by predicating them, but rather that there are too many possibilities – the speaker might mean that the thing is a state of affairs or event that is good for him and his friends, or for people generally, or that it is good for use as an example in a class in philosophy of history, and so on and on.)

So just as there is no such a thing as being just good, there is no such thing as being just a good *K* – for example, just a good book or just a good banana. A fortiori, the instances of being a good *K* are not themselves among the ways of being good.

2.

In some contexts, a person who says 'That was a good act' is likely to mean it was an act that was good to look at; but in other, perhaps more common contexts, a person who says this is likely to mean it was an act that was just, or generous, or kind, or considerate, and so on. An ascription of being just or being generous and so on is praise; and it seems clear that these are further ways in which a thing can be good. But they are intuitively not first-order ways of being good. Certainly, '*X* is just' is not equivalent to anything of the form '*X* is good plus adjunct', and similarly for the rest. The story I am telling says that these are *second-order ways of being good.*

There are other second-order ways of being good besides the moral. There are the aesthetic: being graceful, striking, dazzling, serene, witty, and so on. There are also being charming, elegant, sophisticated, and so on. (Are those aesthetic properties?) What about being healthy? I should think that it, too, is a way of being good. (I should think that being healthy is being in good health.) Is '*X* is healthy' equivalent to anything of the form '*X* is good plus adjunct'? Presumably not, and I therefore include it among the second-order ways of being good.

The context of an assertion of 'That's good' – as of a sentence of the form '*X* is a good *K*' – may make it likely that the speaker means that the thing is good in a first-order way ('good for use in making cheesecake', 'good for Alfred') *or* that it is good in a second-order way ('just', 'graceful').

Why *second-order*? My suspicion is that while none of the ways of being good that I here call second-order is simply reducible to a first-order way of being good, they rest on first-order ways of being good. How? I shall restrict myself to the moral

[7] That lovely example comes from Ziff.

second-order ways of being good, since it is only those which are relevant to the story I am telling.

Two preliminaries. First, it is not merely some acts that are just, generous, and so on: some people are so, too. But it is surely plain enough that we should not take the property of being just to be the "what is in common" to all just entities, people as well as acts, the possession of which marks them all as just. (Compare Moore on goodness in the passage I quoted in note 2 above.)

A helpful model is Aristotle's example of health. Some people are healthy, and so are some foodstuffs; but being healthy is not the property had in common by all healthy things – rather, being healthy is what all healthy *X*s have in common, healthy *Y*s being healthy only derivatively, in the appropriate sense. Surely, the *X*s here are people, and the *Y*s foodstuffs; that is, surely being healthy is what all healthy people have in common, healthy foodstuffs being healthy only derivatively, in the sense that eating them is conducive to a person's becoming or remaining healthy.

So similarly, we may take it that being just is what all just *X*s have in common, just *Y*s being just only derivatively, in the appropriate sense. But which are the *X*s *here*? Some philosophers say that acts are the *X*s: thus they say that being just is what all just acts have in common, just people being just only derivatively, in the sense that they are prone to performing just acts. Other philosophers say that people are the *X*s: thus they say that being just is what all just people have in common, just acts being just only derivatively, in the sense that they are acts that a just person would perform. (I take this to be Aristotle's view.) The notions "just act" and "just person" are presumably interdefinable: which way should analysis proceed? I am myself in the former group; that is, I think that an act's being just is metaphysically prior, and a person's being just metaphysically secondary, and I shall assume that that is so. But I shall not argue for this idea, since the story I am telling does not rely on it. Given interdefinability, the story I am telling could as well have been told the other way round.

In sum, I shall take the noun phrase 'being just' to refer to what all just acts have in common, just people being just only derivatively, in the sense that they are prone to performing just acts.

But I shall follow a common usage according to which the noun 'justice' refers to a character trait possessed by people, namely, the character trait that consists in proneness to performing just acts. (Perhaps it need not be said, though it will be, that this proneness to performing just acts includes proneness to avoiding unjust acts, since it includes proneness to performing just avoidances.)

Similar questions arise in respect of being generous, being kind, and so on.[8] I shall take the noun phrase 'being generous' to refer to what all generous acts have in

[8] It has to be conceded that from this point of view, being just is not as simple a case as the others on the list. Being generous and being kind are possessed only by acts and people. Being just has a wider range of application: some states of affairs and forms of government are also just. Are *they* just only derivatively? And in what sense are they if they are? One could instead argue that it is states of affairs that are

common, generous people being generous only derivatively, in the sense that they are prone to performing generous acts. I shall take the noun ‘generosity’ to refer to the character trait that consists in proneness to performing generous acts. Similarly for ‘being kind’ and ‘kindness’, and so on.

A second preliminary is required because the list of moral second-order properties that I gave trails off into ‘and so on’. What fixes what does and does not belong on the list? I shall have more to say about this in the following section. For the moment, I mean for the list to include all of those act properties *F* such that there is a character trait consisting in proneness to performing *F*-ish acts, and it is a *virtue*. Thus, being just is on the list since there is a character trait consisting in proneness to performing just acts – I am taking the noun ‘justice’ to refer to it – and it is a virtue. For similar reasons, being generous, being kind, and being considerate are also on the list. In light of that condition on membership on the list, I shall call these properties *virtue properties*.

3.

The story I am telling says that the virtue properties are second-order ways of being good, resting on the first-order goodness-for.

How resting? On subjectivist views, for an act to possess a virtue property, it is only intention that matters, and not success. On those views, therefore, an act can possess a virtue property and be good for no one at all, since the world may conspire against the agent and make his just or generous or kind act misfire. I think that a mistake. My own, objectivist, view is that, if a man intends to be acting justly or generously or kindly, that may (or may not) speak well for him personally, but his intention in acting does not at all settle the question whether he has *in fact* acted justly or generously or kindly. I shall not argue the matter here, however.

For even if it is success (and not mere intention) that matters, it is plain enough that an act can possess a virtue property and not be *on balance* good for people, either because it is good for fewer than an available alternative or because it is less good for those affected than an available alternative. The example of Alfred, who is under threat by the Mafia, is of the former kind, that is, good for fewer. The example went like this: if Alfred does not kill Bert, the Mafia will kill three other people. Those who offer this as a countercase to Moore’s story believe that Alfred’s killing Bert would be unjust; they believe that Alfred’s refusing to kill Bert would be just, and that is surely right. Now, Alfred’s refusing to kill Bert would be good for Bert. But it would be bad for the three others. So the fact of the refusal’s being just is entirely compatible with the refusal’s being good for fewer than the available alternative, and thus with its failing to be on balance good for people.

(nonderivatively) just, and everything else (acts, people, governments) at best derivatively just. I must leave these things open. Fortunately, nothing in my story relies on any particular decision about them.

So we cannot say that an act possesses a virtue property only if it is on balance good for people. If the virtue properties rest on goodness-for, then they do so in a more complex way than that.

I cannot say now exactly what that more complex way is. There is a gap in my story here. Still, there is a familiar idea about the virtues that I think must surely be right, and if it is, then the following may anyway point toward a way of filling the gap.

What I have in mind is the idea that the fact of there being people who possess the virtues is good for us. On some views something stronger is the case, namely, that a person's possessing a virtue is good for that very person. Perhaps that stronger idea is correct. Even if not, however, there being people who possess the virtues is plainly good for the rest of us.

Justice is an obvious case. Justice is proneness to performing just acts; that is, it is proneness to doing what one *owes* to others – whether in the way of keeping one's word, or refraining from taking advantage of the weakness or ignorance of others, or carrying one's fair share of the community's burdens, and so on. No doubt a particular just act may not be on balance good for people (as in the example of Alfred). But it is better for us that the people among whom we live be just than that they not be just. Indeed, this is not merely better for us, but essential to us, since we can form a community at all – and thereby obtain benefits which are essential to us and which only community can provide – only if a substantial number of those among whom we live are just.

I stress: whatever else may be true of the people among whom we live, it is better for us that they be just than that they not be just. Even if they are not generous, not kind, and not considerate. For it would be worse for us if they were also not just.

What I suggest now is that we should take this to be, not merely a fact about justice, but what marks it as a virtue.

Is honesty a virtue? Well, what exactly is it? If honesty is (as some people think it is) a sheer unwillingness to lie, come what may, then on my view it should not turn out to be a virtue. (Indeed, on my view it is a peculiarly unattractive form of self-righteousness, and thus a minor vice.) And it does not turn out to be a virtue on this account of the matter, for it is better for us if the people among whom we live are ready to lie when they do not owe it to their hearers not to lie and lives will otherwise be lost. If honesty is, more narrowly, an unwillingness to lie where one does owe it to one's hearers not to lie, then it is a virtue, a subvirtue falling under justice.

Is prudence a virtue? Again: it depends on what prudence is. If prudence is (as some people think it is) entirely self-advantaging, if, that is, it is carefulness only where one's acts may cause harm to oneself, then prudence should not turn out to be a virtue, and does not on this account of the matter. If prudence includes, more broadly, carefulness where one's acts may cause harm to others, then it should turn out to be a virtue. But then so understood, it, too, is a subvirtue falling under justice.

For obvious reasons, we can call this nested cluster of virtues the *reliance virtues*.

There is a second cluster of virtues, which we can call the *virtues of concern*. They include generosity, kindness, and considerateness, and they have two features, the first of which marks them as virtues, the second as virtues distinct from the reliance virtues. The first is that they are pronenesses to doing what is good for others at a cost, at most, to their possessors. (It is not generosity in me to take from you to give to others. On the other hand, I say 'at most' since a kind act may be entirely costless; by contrast, an act is not generous unless it is in a measure costly to its agent.) In light of their having this first feature, they have a direct connection with goodness-for, more direct than the reliance virtues. It is plain that whatever else may be true of the people among whom we live, it is better for us that they possess these traits than that they do not.

Their second feature is that the acts these virtues are pronenesses to performing are not called for by justice; that is, it is not unjust to omit them. (However much good it may do my grocer for me to pay his bill, paying it is not an exercise of generosity or kindness or considerateness.) In light of their having this second feature, they are not among the reliance virtues: a person can be just without possessing any of them.

A person might have a proneness to "maximizing goodness-for" – that is, he is prone to doing a thing wherever it would be on balance better for people that he do it than that he not do it. That proneness also has a direct connection with goodness-for. But two things should be noticed about it. On the one hand, it is not a virtue of concern, since it fails to meet the first of the two conditions I mentioned, for when I do what I owe it to you not to do, because it would be on balance better for others that I do it, then my act imposes a cost (not on me, but) on you. On the other hand, while it might have seemed, off the cuff, to be better for us that those among whom we live possess this trait than that they do not, that is not true. If we cannot count on others' keeping their word to us unless we are so far lucky as that their keeping their word to us maximizes goodness-for, then we (simply) cannot count on others' keeping their word to us. The bearing of this on the possibility of our forming a community is obvious.

Perhaps the virtues of concern nest in the way in which the reliance virtues do, with generosity dominating them as justice dominates the reliance virtues. No matter, for present purposes.

What about courage? Courage appears on many philosophers' lists of the virtues, but I have so far postponed discussion of it. Suppose we believe that (i) courage is a virtue. What is it? Suppose we believe, as I do, that (ii) courage is (roughly) steadfastness of purpose, proneness to "standing firm," even in face of danger. This says nothing about what the courageous person's purposes are – it leaves open what they are – and that may seem to make trouble for my test for being a virtue, which says that (iii) a trait is a virtue just in case whatever else may be true of the people among whom we live, it is better for us that they have the trait than that they not have it. For if the people among whom we live are just, then all is well, it is better

for us that they also be courageous. But if they are unjust, then it is (much!) better for us that they be cowards. In sum, if (ii) and (iii) are true, then courage is not a virtue. But (i) says it is a virtue, so something has to give.[9]

I suggest that it is (i) that should give. Courage subserves justice in the sense that just acts require at least a minimum measure of courage. (It also subserves generosity, for generous acts may also require courage, and very generous acts may require great courage.) But it also subserves injustice, for unjust acts also require at least a minimum measure of courage, and very unjust acts may require great courage. It is hard to see how a trait that subserves a major vice can plausibly be thought of as, itself, a virtue.

Courage is not unique in this respect. Compare industriousness. (What is this industrious man industrious *at*?) A minimum measure of industriousness is required by just acts; but very unjust acts may require it, too. Again, compare loyalty. (Who or what is this loyal man loyal *to*?) A minimum measure of loyalty is required by just acts; but very unjust acts performed by groups are likely to require it, too. Similarly for conscientiousness. (What are the contents of this conscientious man's conscience?)

In sum, then, I suggest that we should not include courage – or industriousness, loyalty, or conscientiousness – among the virtues. Let us return now to what I was calling *virtue properties.* I said I meant for the list to include all those act properties *F* such that there is a character trait consisting in proneness to performing *F*-ish acts, and it is a virtue. Justice and generosity are pronenesses to performing just and generous acts, and they are virtues; so being just and being generous are virtue properties. Courage is not a virtue; hence being courageous is not a virtue property.

That, I think, is as it should be. The properties I am calling virtue properties are those I indicated at the beginning of section II: being just, being generous, being kind, being considerate, and so on. I said that an ascription of these is praise, and that it seems clear that these are further ways of being good – that is, ways of being good in addition to the first-order ways of being good. Is an ascription of being courageous praise? When we ascribe being courageous we typically are praising; that is because we do not typically ascribe being courageous where we think ill of the act in question on other grounds. (Nobody says of a particularly villainous act, "Well, it was good in one way anyway: it was courageous." Nobody for a moment considers giving medals for courage to courageous villains.) These facts can be explained if we take seriously that an act's being courageous just is its having been performed despite danger to its agent, for this suggests that the praise goes to the act for occurring at all. Suppose a man risked his life to save another. That was a very generous act. It was also very courageous. To praise it on the former ground is to praise it for what it was; to praise it on the latter ground is to praise its occurring

[9] If we accept (ii) we are committed to accepting that it is entirely possible for a person to be both courageous and a villain. Geach and Foot reject this possibility; see Geach, *The Virtues* (New York: Cambridge, 1977); and Foot, "Virtues and Vices," in her *Virtues and Vices* (Cambridge: Blackwell, 1978). But I have been unable to become clear about what alternative construal(s) of courage they would have us adopt.

despite the danger to its agent. That is why praise of an act on ground of its being courageous is parasitic on there being other grounds for welcoming the act. (It also explains why no praise goes to courageous acts of villainy.)

All of this is rough, however, and I must leave it open. As I said, there is a gap in my story here. I have wished merely to point toward what *may* be a way of filling it.

4.

What we have so far is this. The first-order ways of being good are the likes of being good for use in, good at, good for, and so on. There are also second-order ways of being good: among others, the virtue properties being just, generous, kind, and so on. These rest on a particular first-order way of being good, namely, goodness-for. (Exactly how do they rest on goodness-for? I made a suggestion, but it was no more than that.)

If what we have so far is correct, then we are home free: the rest of the story is briefly told, for we can get to the right relatively easily from the virtue properties. In short, what we are morally required to do is to avoid their contraries. Morality requires us to *do* a thing if and only if not doing it would be unjust, or mean, or cruel, and so on. Morality requires us not to do a thing if and only if doing it would be unjust, or mean, or cruel, and so on.[10]

Some philosophers have been saying in recent years that moral philosophers should look at the virtues, and some among them have said, more strongly, that moral philosophers should not only look at the virtues, they should also overcome their fixation with moral requirement. I agree that it would be profitable to look at the virtues. But that is because of the importance of their contraries, the vices, to supplying an account of what morality requires us to do.

My way of putting the matter might seem more complicated than it need have been. Why put it in the negative rather than the positive? Why say we are morally required to avoid injustice, meanness, and cruelty instead of saying that we are morally required to act justly, generously, and kindly? Moral requirement seems to me to be a rather weak notion: the standard it imposes is that of mere human decency. Now, generosity, kindness, indeed justice itself, can be very costly, and there are cases in which paying those costs would be supererogatory. Morality may require us to pay those costs, but I suggest that it does so only where refusing to pay them would be mean, cruel, or unjust.

[10] For my own part, I believe we should accept an objectivist view of the requirements of morality, that is, a view according to which a person's having done what morality requires him to do turns on success rather than on intentions – just as, for my own part, I believe we should accept an objectivist view of the virtue properties (and their contraries), that is, a view according to which an act's possessing a virtue property (or its contrary) turns on success rather than on intentions. (See the second paragraph of section III above. I regret having wobbled about this matter in *Moral Relativism and Moral Objectivity*.) I do not argue for these things here, however. What is surely plain, and is in any case required by this part of my story, is that we should be either objectivist about both or subjectivist about both.

To return just briefly, then, to Alfred and the Mafia: I said it is hard to see what could make Alfred's killing Bert wrong if it is not bad, and then hard to see how the Mafia's killing the other three people could fail to be three times as bad. Now we can say: the way in which Alfred's killing Bert would be bad is that it would be unjust. The way in which the Mafia's killing the three others would be bad is that it, too, would be unjust. Morality, therefore, requires both Alfred and the Mafia to refrain. That they will not refrain from killing their three if he refrains from killing his one leaves that conclusion entirely unaffected.

This solution to the problem relies on one's willingness to agree that the fact that the Mafia will not refrain from killing their three if Alfred refrains from killing his one does not mean that Alfred owes it to the three to save them from the Mafia. It seems to me plain that that fact does not mean this. Fortunately for my story, there is reason to think that Alfred's not owing it to the three to save them can be explained by appeal to the first-order goodness-for. A moral regime under which I can make you owe people something simply by threatening not to pay my own debts to them is a moral regime with a massive free-rider problem, and thus is a regime that would be bad for all of us.

Two points remain to be taken note of. First, a potential difficulty for my account of moral requirement emerges as follows. Suppose we believe that the following is possible:

(1) Bert's not doing such and such would be unjust, but Bert's doing the such and such would be cruel.

Then my account of moral requirement yields that the following is possible:

(2) Morality requires Bert to do such and such and morality requires Bert not to do it.

Is that possible?

On some views, (2) is entirely possible.[11] If it is, then the potential difficulty is not an actual difficulty.

But suppose we believe (as I do) that (2) is not possible. Then we may seem to have an actual difficulty, for either we must give up the account of moral requirement that I have offered, or we must agree that (1) is not possible. Which to do?

It will not surprise that I suggest we should agree that (1) is not possible. I have no argument to the effect that it is not. But I simply cannot imagine how the following could be true: justice in a given case calls for doing what it is cruel to do. Justice may, of course, require harming people, as, for example, where it calls for imposing a severe punishment; proceeding, where it does, is not cruelty. (What is cruel is hurting or harming gratuitously.) Similarly, I believe, for any other pair of

---

[11] See the literature on "moral dilemmas" – for example, *Moral Dilemmas and Moral Theory*, H. E. Mason, ed. (New York: Oxford, 1996).

contraries of virtue properties. So for my part, too, the potential difficulty here is not an actual difficulty. But others may think otherwise.

The second point that remains to be taken note of is that my account of moral requirement construes it as entirely other-regarding. That is a consequence of the fact that my account of the virtues, and therefore my procedure for picking out the virtue properties, construes them as fundamentally social: that is, my account of what fixes whether a character trait is a virtue, is its effect on others. But that a story about the right and the good yields that moral requirement is other-regarding seems to me exactly as it should be. Gluttony, fecklessness about one's own interests, and excessive timidity are a bad business; they are certainly bad for their possessor, but on my view they are no breach of moral requirement – unless they issue in injustice, meanness, cruelty, and the like. That is because, as I said, moral requirement seems to me to be a rather weak notion. I said that the standard it imposes is that of mere human decency, and what I meant was human decency in dealings with others.

5.

What precedes is two thirds of the story about the right and the good I find attractive. In sum, I began with the first-order ways of being good, and went up from there to the second-order ways of being good that consist in the virtue properties, and then up from their contraries to the requirements of morality. We need now to return to the first-order ways of being good, and go down from there.

What I mean becomes clear if we return to Moore's story for a moment. I said that Moore's story begins with the good. Some things are good, Moore said, and some things are not; so there is such a property as goodness. I said that Moore had much to say about which things do and which things do not possess the property goodness, but that I would not summarize that part of his story. My reason should now be clear: there is no such property as goodness. Now, anyone who likes the story I have been telling so far is going to need to say something about which things are good in the first-order ways. What makes a thing be good to look at? What makes a person be good at hanging wallpaper? What makes a thing be good for use in making cheesecake? In one respect, these questions seem even more pressing than their analogue in the case of Moore's story, for surely the first-order ways of being good are not a mere happenstance clutter. There must be something they have in common that marks them all as ways of being *good.*

Expressed in the formal mode: it is not mere happenstance that the word 'good' appears in the expressions 'good to look at', 'good at hanging wallpaper', 'good for use in making cheesecake', and so on: its meaning makes a contribution to their meanings, and we need to see what that contribution is.

Intuitively, for a thing *X* to be good in one of the first-order ways is for *X* to *benefit* someone or some thing *Y* (which might or might not be *X* itself) in the appropriate way, or to be capable of doing so. There has to be something "in it" for some *Y*, a gain or potential gain of the appropriate kind, if *X* is to be good in a first-

order way. That, my story says, is what the first-order ways of being good have in common.

What are the ways in which *X* might benefit *Y*? *X* might please *Y*. *X* might answer to *Y*'s wants. There are other possibilities, too, one of which will emerge as important in the following section.

Take being good to look at, taste, or listen to. A thing that has one or other of those properties has them in that it pleases people: looking at it pleases, tasting it pleases, listening to it pleases, and so on.

Take being good at doing something, as it might be, hanging wallpaper or playing chess. A person who is good at hanging wallpaper is one who is capable of hanging wallpaper as people typically want that done when they want it done. (There is nothing in the nature of wallpaper that fixes how it should be hung; there is only how people typically want it hung.) A person who is good at playing chess is one who is capable of winning when playing with experienced players, thus is capable of answering to his own wants when playing with them. (Unlike wallpaper hanging, chess has rules that fix what counts as winning, and you are not playing chess on a given occasion – you are only dabbling, playing at chess – if you are not then trying to win.)

Take being good for use in doing this or that. Spring-form pans are good for use in making cheesecake: that is because getting a cheesecake out of an ordinary pan is hard, and results in a mangled cake, and people who want to make cheesecake typically want the making of it to be easy, and the result to be smooth and round.

Moreover, where what is wanted varies from place to place, so does what is good for use in doing this or that. In Norway, lye is good for use in preparing codfish; that is a function of Norwegian wants – not merely of what pleases them but also of what is and is not cheaply available there. Not so in America. (It would be weird to say that lye is really good for use in preparing codfish everywhere, it is just that Americans do not know that. It would be equally weird to say that lye is not good for use in preparing codfish anywhere, it is just that Norwegians do not know that.)

I have so far mentioned three of the first-order ways of being good: first, being good to look at, taste, listen to; second, being good at doing a thing; third, being good for use in doing a thing.[12] Let us focus on a fourth, namely, being good for a thing. That way of being good is crucial to my story since it is what the story says the virtue properties rest on.

6.

Very many inanimate objects *Y* are such that for some *X, X* is or would be good for *Y.* I begin with artifacts. Regular oiling is good for a lawn mower. How so?

All artifacts have one or more of what might be called *design functions,* where 'Among the design functions of *Y* is to *A*' is true just in case *A*-ing is among the

[12] For more detail, see *Moral Relativism and Moral Objectivity*. I should add, however, that the account of goodness-for which appears in the following sections is different from the one I offered there.

things *Y* is designed to do.[13]

Where *Y* is an artifact, *Y* is designed to *A* in that it was designed by a human designer to *A*. That is, a human designer selected some features *F* for *Y* because *Y*'s having *F* would increase the likelihood that it would *A*. Thus, among the design functions of any lawn mower is to enable its user to mow lawns with it, since some of its features were selected for it because its having them would increase the likelihood of its enabling its user to mow lawns with it. Another of the design functions of a lawn mower is likely to be to enable its user to mow lawns safely with it, since some of its features are likely to have been selected for it because its having them would increase the likelihood of its enabling its user to mow lawns safely. (I doubt that any artifacts have only one design function. Whatever the use to which a manufactured item is to be put, some of its features are likely to have been selected, not just to increase the likelihood of its being usable for that purpose, but also to increase the likelihood of its being safe to use for that purpose.)

Now, I said in the preceding section that intuitively, for a thing *X* to be good in one of the first-order ways is for *X* to benefit someone or some thing *Y* in the appropriate way, or to be capable of doing so. I suggest, then, that we should say: if *Y* is an artifact, then (i) *X* is good for *Y* just in case *X* benefits *Y*, and (ii) *X* benefits *Y* by and only by conducing to *Y*'s doing what it is among *Y*'s design functions to do. For example, regular oiling is good for a lawn mower and that is a consequence of the fact that regular oiling benefits it, which it does by conducing to the lawn mower's doing what it is among its design functions to do, namely, to enable its user to mow lawns with it.

What is important is that on this account of the matter, the benefit supplied by *X* does not consist in anybody's being pleased, or in anybody's wants being answered to: the benefit supplied by *X* is to the artifact *Y* itself. That seems to me as it should be. Regular oiling is good for a lawn mower, and that is not because regular oiling of a lawn mower pleases anybody, or answers to anybody's wants. No doubt it was the wants of consumers, and therefore the wants of manufacturers, that fixed that lawn mowers were designed in such a way as to have, among other design functions, the design function of enabling their users to mow lawns. No doubt people buy lawn mowers nowadays because they want to be able to mow their lawns, and are pleased when they are able to do so. But regular oiling would still be good for a lawn mower even if people stopped wanting to mow their lawns, and even if manufacturers therefore stopped making lawn mowers. Suppose that starting next year, lawn mowers pile up in garages and warehouses, rusting away from lack of care because of lack of users. Oiling a lawn mower would not benefit any person, but would all the same benefit *the lawn mower*; and oiling it would benefit it precisely because, though nobody wants to use it to mow lawns, oiling it would conduce to its being in

---

[13] I adapt this notion *design function* from what is currently the most widely accepted analysis of functions; I have been particularly helped by Philip Kitcher, "Function and Design," *Midwest Studies in Philosophy*, XVIII (1993): 379-97. (Design functions are different from what I shall later call *use functions* and *role functions*.)

a condition in which it *could* be used to mow lawns. That, after all, was what it was designed to do.

We can reput the matter in the following way. Many things have one or more of what might be called *use functions,* where 'Among the use functions of *Y* is to *A*' is true just in case some people do in fact use *Y* to *A*. I should think that as things are, the use function of all (working) lawn mowers is to mow lawns, since some people do use them to mow lawns. But two facts remain. (a) The design functions of a lawn mower remain what they are no matter what its use functions come to be. Thus, the design functions of a lawn mower continue to include 'to enable its user to mow lawns with it', even if the lawn mower – indeed, even if all existing lawn mowers – cease to have any use function (as by being allowed to rust in warehouses) or acquire a novel use function (as by coming to be used merely as garden ornaments). (b) Where a lawn mower's use functions diverge from its design functions, what is good for it is not conducing to its doing what it is among its use functions to do, but rather conducing to its doing what it is among its design functions to do.

Similar points hold about some animate things, which I shall turn to in the following section.

Let us first look briefly at inanimate objects that are not artifacts. If *Y* is inanimate but not an artifact, *Y* has no design functions. Yet it is possible that something, *X*, is or would be good for it. It is plausible to think that, if *X* is good for such a *Y*, then that is because *X* conduces to *Y*'s being in a condition that answers to people's wants. Where people have conflicting wants in respect of *Y*'s condition, it may (or may not) be easy to say whose wants count. Imposing regulations on disposal of wastes in the Charles River in Massachusetts was good for it. How so? It conduced to the river's being in a condition that answered to the wants of those who wanted to fish in it, swim in it, sail small boats on it, and so on. Why is it their wants that fixed what was good for the river? Why not the wants of those who wanted to dump wastes in it? I am going to bypass this question. For present purposes, what matters is only that it is people's wants that fix what is good for it. If the condition of the Charles stopped mattering to people – and were never again going to matter to anybody – then nothing at all would any longer be good (or bad) for it. Rivers are like wallpaper and unlike lawn mowers: just as there is nothing in the nature of wallpaper that fixes how it should be hung, so also there is nothing in the nature of rivers that fixes what condition they should be in or come to be in.

7.

All animate objects *Y* are such that for some *X*, *X* is or would be good for *Y*. Let us begin with plants. Watering a certain plant might be good for it. How so?[14]

---

[14] It has been suggested to me that the answer is simple: watering the plant would be good for it just in case, and because, it needs water. And, quite generally (and in particular, for artifacts as well as plants), that

(1) Getting alpha would be good for *Y*.

Plants, too, are designed to do things, and thus they, too, have design functions. Where *Y* is a plant, *Y* is designed to *A* if it was designed by nature to *A*. That is so if *Y* has some features *F* because possession of *F* by *Y*'s ancestors increased the likelihood that they would *A*, where increasing the likelihood that they would *A* conduced to their reproductive success.[15] Thus, among *Y*'s design functions might be to attract pollinators of kind *K*; that is so if *Y* (as it might be) flowers at a certain time of the year because its ancestors' flowering at that time of year increased the likelihood that they would attract pollinators of kind *K*, and doing so conduced to their reproductive success. Another of *Y*'s design functions might be to protect itself from disease; that is so if *Y* (as it might be) grows a scab or scar when injured because its ancestors' growing a scab or scar when injured increased the likelihood that they would protect themselves from disease, and doing so conduced to their reproductive success.

Indeed, I should think that every animate object *Y* has the following design function: to survive. (More precisely, no doubt: to survive to a time, if any, after which further survival would impede reproductive success.) For every animate *Y* has some features *F* because its ancestors' possession of *F* increased the likelihood that they would survive, and doing so conduced to their reproductive success.

I suggest, then, that we should say about plants what I suggested we should say about artifacts: if *Y* is a plant, then (i) *X* is good for *Y* just in case *X* benefits *Y*, and (ii) *X* benefits *Y* by and only by conducing to *Y*'s doing what it is among *Y*'s design functions to do. In particular, watering a certain plant is good for it if watering it conduces to its surviving.[16]

---

is analyzable into

(2) *Y* needs to get alpha.

This will not do, for (1) and (2) are not even equivalent. What is equivalent to (2) is surely:

(3) Not getting alpha would be bad for *Y*.

and (1) is not equivalent to (3). (For very many, if not all, things *Y*, there are things alpha such that getting alpha would be good for *Y* though not getting alpha would not be bad for *Y* – and those are therefore things which *Y* does not really need.)

It may be suggested that we should say, "Well, anyway, (3) is analyzable into (2)." Why say that rather than that (2) is analyzable into (3)? In fact, if we are unclear about what makes one of them true, we are not helped by being told that what makes it true is what makes the other true – any source of unclarity about one is equally a source of unclarity about the other.

[15] "[O]ne of Darwin's important discoveries is that we can think of design without a designer" – Kitcher, p. 380. Many people, including Kitcher, recommend taking *Y*'s recent ancestors to be what matter. I bypass all the important questions that arise here.

[16] Some people say that what conduces to a plant's survival is good for it, but that what conduces to a plant's reproducing is good only for its species, and not for the plant itself. Why should we believe this? (Can those who say it harbor the idea that, like many human beings, plants want to survive but do not care about reproducing?) A further difficulty is that it is by no means clear what could count as something's being good for a species. is it bad for a species to become extinct? No doubt it might be bad for us that a species become extinct: What could possibly make it bad for the species itself to do so?

I am grateful to Paul Horwich for drawing my attention to the fact that many people say that a species' becoming extinct would be bad – not bad for us, not bad for it, but just plain a *bad thing*. I have said that there are no such properties as goodness and badness; so what can those people be taken to mean? We should note, first, that nobody says this about just any species. I gather that hundreds of

So just as in the case of artifacts, what fixes what is good for a plant is not what pleases anybody, or what answers to anybody's wants. Feeding the dandelions in my front lawn a high nitrogen fertilizer would be good for them even though what *I* want is that they die, before reproducing if possible. A plant's use functions, if it has use functions, are irrelevant, except in so far as their use functions converge with their design functions.

8.

People are obviously a more complicated affair than plants. They, too, have design functions: there are many things they are designed to do since they were designed by nature to do them. Among their design functions is to protect themselves from disease: like plants, they grow a scab or scar when injured because their ancestors' doing this increased the likelihood that they would protect themselves from disease, and doing so conduced to their reproductive success. Another of their design functions is to survive.

But there are also things they are designed to do that they were not designed by nature to do, but rather by themselves. Just as a human designer can select some features *F* for an artifact because the artifact's having *F* will increase the likelihood that it will *A*, so also can human beings select some features *F* for themselves because their having *F* will increase the likelihood that they themselves will *A*. For example, a man may select unusually strong muscles for himself (going into training to acquire them) because his having them will increase the likelihood that he will win at weight lifting.

Some people are happy to accept the following sufficient condition for use of the term 'goals' (and analogous conditions for 'aims' and 'purposes'):

(1) *Y* has among its goals to *A* if *Y* has among its design functions to *A*.

---

species of ants (or is it termites?) become extinct every day: is anyone seriously inclined to call that just plain a bad thing? Pandas are another matter, however. Dear living teddy bears! But that points to the way in which their becoming extinct would be bad: we would lose something if they became extinct, and thus their becoming extinct would be bad for us. (There are other losses besides losses of the useful. Consider losing your old childhood nonliving teddy bear. Or your child.) So far as I can see, it is in being bad for us, and only in being bad for us, that their becoming extinct would be bad.

The species "human being" is yet another matter. If humans become extinct because of something that causes the last of them pain, then that would be bad for them. Suppose they become extinct because the last of them want not to reproduce, and that that want is "ideal" in a sense to emerge in the text below. (Not due to ignorance, and so on.) Thereafter the world would go rolling on to its end without any human intelligence in it, without any high intelligence in it at all unless some other highly intelligent species arises. Would that be bad? (In what way?) Is the sadness that prospect provokes in some people like the sadness that some people feel at the prospect of their dying without issue? (*Is* dying without issue bad? In what way?) I leave these dark questions open.

It is a consequence of (1) that plants and lawn mowers as well as people have goals; the people I refer to are happy to accept that consequence. I am sure that at least some among them also accept:

> (2) *Y* has among its goals to *A* if *Y* chooses behaviors it thinks will increase the likelihood of its *A*-ing, and does so because it wants to *A*.

It is not a consequence of (2) that plants and lawn mowers have goals. But some things meet both sufficient conditions. My weight lifter has among his design functions to win at weight lifting, and so has that as his goal under (1); he also chooses behaviors he thinks will increase the likelihood of his winning at weight lifting, and does so because that is what he wants to do, and so has that as his goal under (2).

On the other side are people who reject (1), and opt for a strengthening of (2), namely:

> (3) *Y* has among its goals to *A if and only if Y* chooses behaviors it thinks will increase the likelihood of its *A*-ing, and does so because it wants to *A*.

According to (3), only people and higher animals have goals.

There is no need for us to decide which side is correct. I shall accept (3), but largely because doing so provides a perspicuous way of abbreviating the theory I think we should accept.

It is a familiar theory. It says, in short: if *Y* is an adult person, say Alfred, then (i) *X* is good for Alfred just in case *X* benefits Alfred, and (ii) *X* benefits Alfred by and only by conducing to his reaching one or more of his goals. (It may pay to make explicit that I mean to include enabling him to reach his goals under 'conducing to' his reaching his goals.)

But that is too short, for not just any goal Alfred happens to have counts for the purposes of this theory: a goal counts only if it meets two kinds of constraints: (a) Alfred does not have the goal because of ignorance of, or lack of careful attention to, the costs of doing *A*, including opportunity costs, and (b) Alfred does not have the goal because of such improper "preference bendings" as threats, drugs, or a grossly unjust political regime.

A familiar way of accommodating these constraints is by reputting the matter counterfactually: what counts is what Alfred would aim at in ideal conditions of full information about costs, assessed "in a cool hour," and lack of improper preference bendings.

Why do those constraints have to be imposed? Alternatively put, why does it matter that Alfred would aim at this or that in ideal conditions, given he does not in fact aim at this or that? The answer is that what is in question here is what *really*

benefits Alfred, not what he may happen to think would benefit him – off the cuff, in ignorance, under threat, and so on.

In sum, then, X is good for Alfred just in case X benefits him in that it conduces to his reaching one or more of (what I shall call) his *ideal goals.*

I said that this is a familiar theory, but it is certainly not universally accepted. Might it not turn out that Alfred's ideal goals are such that reaching them precludes his having family and friends, or a wide variety of intellectual and aesthetic enjoyments, or even mere good health? Would it not benefit him to conduce to his life's including those things, despite his ideal goals? Well, would it? Proponents of the familiar theory reply: if we think it would, then that is because having those things in our own lives is important to us, but it is mere hubris to suppose that getting them benefits every person, whatever his ideal goals may be, and however drastically making room for them in his life might interfere with his chances of reaching his own ideal goals. Who are we to tell the saint or research scientist or great chess master to make some friends, read some good novels, listen to some good music, keep out of unhealthy environments – assuming that his actual goals are his ideal goals?

What if Alfred's ideal goals include or require doing harm to others? Proponents of the familiar theory reply that while it *may* be true that what benefits one does no harm to others, this is by no means an a priori truth.

These replies seem to me to be right. So the familiar theory also seems to me to be right.

Some things that follow from that theory might be worth drawing attention to. In the first place, Alfred's use functions, if he has any, are irrelevant to what is good for him; what is good for him is not fixed by what benefits others who use him for this or that purpose but by what benefits him.

Perhaps people do not on the whole have use functions. But they mostly (all?) do have what might be called *role functions.* If Alfred's job is to keep the books at the local shoestore, then keeping those books is among his role functions. Fine. But while conducing to his keeping those books may be good for him, it also may not be.

A third and more interesting consequence of the familiar theory is this: not only do use functions and role functions fail to fix what is good for a person, but so also do design functions. Suppose that Alfred was designed by nature to A. Conducing to his A-ing may conduce to his reaching one or more of his ideal goals. But it may not. Doing so may even conflict with his reaching one or more of his ideal goals. If conducing to his A-ing does not conduce to his reaching an ideal goal, then the familiar theory tells us that doing so is not good for him. So design functions have not the role in fixing what is good for a person that they have in fixing what is good for artifacts and plants.

But this, too, seems to me to be right. It is among Alfred's (naturally selected) design functions to survive. But the hero or saint may try to save the life of another at risk to his own life. It cannot at all plausibly be thought to be good for him to prevent him from doing so.

Again, it is among Alfred's (naturally selected) design functions to be able to engage in such physical activities as running. It is hard to imagine a person whose ideal goals are so eccentric that an ability to run would not in any way help him to reach them. But suppose that is true of Alfred: suppose he cares not the least about being able to run. Then conducing to his being able to do so is not good for him. That seems to me very plausible.

Things are otherwise when it comes to Alfred's body parts. It is among the (naturally selected) design functions of Alfred's legs to enable Alfred to run. Legs have no goals, however, so what is good for them is fixed by what conduces to their doing what it is among their design functions to do. Alfred's ideal goals are conclusive when it comes to what is good for him; they are irrelevant to what is good for his legs, just as they are irrelevant to what is good for his lawn mower and his plants.

9.

In sum, the story I have been telling went as follows. I began by saying that there is no such property as goodness; there are rather being good for use in, being good at, being good to, being good for, and the like. I called these first-order ways of being good. The second-order ways of being good include the virtue properties: being just, generous, courageous, kind, and so on. These, I suggested, rest on the first-order goodness-for. From them we can get to moral requirement by way of the following thesis: what we are morally required to do is to avoid their contraries. If these ideas are correct, then the right does rest on the good, though unfortunately not in the very simple way described in Moore's story.

I suggested also that the first-order ways of being good rest on benefiting in appropriate ways, which may involve pleasing, or answering to wants, or conducing to something's doing what it is its design function or ideal goal to do. This may well be the most controversial part of the story. It should be stressed, however, that the rest of the story is independent of this part. What the fact of this part's being unacceptable would mean (if it is unacceptable) is just that some better account of the expressions by which we attribute the first-order ways of being good had better be found. For it is not just happenstance that the word 'good' appears in all of those expressions: its meaning surely does contribute to their meanings, and we need an answer to the question how it does.

CHAPTER 12

M. J. ZIMMERMAN

DEFENDING THE CONCEPT OF INTRINSIC VALUE

Despite the initially favorable reception enjoyed by *Principia Ethica*, much of what Moore had to say about intrinsic value in that book was, and continues to be, strongly disputed. The view that goodness is a simple, nonnatural property has been criticized by many people and in many ways. Some have argued that goodness is an analyzable property[1] (something I'll discuss in detail in Chapter 4). Others have argued that it is a natural property (or relation)[2] (something I'll discuss only incidentally). Still others have argued, more radically, that goodness is not a property (or relation) at all[3] (something I will not discuss). But none of these critics have rejected the very idea of goodness. None of them, that is, have contended that to say of something that it is good (in Moore's sense) is to speak nonsense. Yet just this is the charge of certain recent critics. Peter Geach was one of the first of these critics.[4] Bernard Williams has endorsed these criticisms.[5] Philippa Foot has also indicated that she rejects the notion of goodness that consequentialists such as Moore invoke.[6] Finally, and most recently, Judith Thomson has in a number of places forcefully pressed the view that there is no such thing as the sort of goodness with which Moore is concerned.[7]

Obviously, if these important authors are right in what they say, then my project in this book is ill conceived and I have wasted my time writing it (and you are wasting

[1] See, for example, A. C. Ewing, *The Definition of Good* (New York: Macmillan, 1948).
[2] See, for example, R. B. Perry, *General Theory of Value* (New York: Longmans, Green, 1926).
[3] See, for example, noncognitivists such as A. J. Ayer, Charles Stevenson, and R. M. Hare.
[4] Peter Geach, "Good and Evil," *Analysis* 17 (1956): 33-42.
[5] Bernard Williams, *Morality* (New York: Harper and Row, 1972).
[6] Philippa Foot, "Utilitarianism and the Virtues," *Mind* 94 (1985): 196-209.
[7] Judith Jarvis Thomson, "On Some Ways in Which a Thing Can Be Good," in *The Good Life and the Human Good*, ed. E. F. Paul et al. (Cambridge: Cambridge University Press, 1992); Thomson, "Goodness and Utilitarianism," *Proceedings and Addresses of the American Philosophical Association* 67 (1994): 7-21; Thomson, "Moral Objectivity," in *Moral Realism and Moral Objectivity*, by Gilbert Harman and Judith Jarvis Thomson (Oxford: Blackwell, 1996); and Thomson, "The Right and the Good," *Journal of Philosophy* 94 (1997): 273-98 [* pp. 131-52 of this volume].

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 153-168.
© 2005 *Springer. Printed in the Netherlands.*

your time reading it). So I had better try to refute their attack on the concept of intrinsic value. Since Thomson's attack is the most detailed (and the most recent), I'll pay particularly close attention to it.[8] But, as she builds upon Geach's observations, I'll begin by considering what he has to say on this issue.

1. "GOOD" AND "YELLOW": A FLAWED ANALOGY

We can draw a grammatical distinction between two ways in which adjectives can operate. An illustration will help make the distinction clear. "Red" operates as a grammatically *attributive* adjective in the phrase "a red book" and as a grammatically *predicative* adjective in the phrase "this book is red." Geach employs this same terminology to make a related logical distinction. As he puts it:

> I shall say that in a phrase "an A B" ("A" being an adjective and "B" being a noun) "A" is a (logically) predicative adjective if the predication "is an A B" splits up logically into a pair of predications "is a B" and "is A"; otherwise I shall say that "A" is a (logically) attributive adjective.[9]

Thus "red" is itself said by Geach to be a logically predicative adjective, since ***x* is a red book** "splits up logically" into ***x* is a book** and ***x* is red**. But other adjectives don't sanction such "splitting up" and so must be said to be logically attributive. Geach gives these examples: ***x* is a big flea**; ***x* is a small elephant**; ***x* is a forged banknote**; and ***x* is the putative father of *y***. In the course of his remarks, Geach also suggests a second, related test by means of which we can distinguish between logically predicative and attributive adjectives.[10] Logically predicative adjectives sanction what I'll call "transposition," whereas logically attributive adjectives do not. For example, from ***x* is a red bird** and **a bird is an animal** we may infer ***x* is a red animal**. With "big," "small," "forged," and "putative," however, no such inference is warranted.

In *Principia Ethica*, Moore likened "good" to "yellow," declaring that both express simple properties, the latter a natural property, the former a nonnatural one.[11] But Geach claims that the analogy is badly flawed because, whereas "yellow" is, like "red," a logically predicative adjective, "good" is logically attributive. This is because "good" sanctions neither splitting up nor transposition. For example, ***x* is a good car** cannot be split up into ***x* is a car** and ***x* is good**. Also, from ***x* is a good cricketer** and **a cricketer is a person** we may not infer ***x* is a good person**. Geach concludes:

> Even when "good" or "bad" stands by itself as a predicate, and is thus grammatically predicative, some substantive has to be understood; there is no

---

[8] I will focus on Thomson (1997), her latest presentation of these claims; but I will also draw on the other works cited in the last note, especially Thomson (1992).
[9] Geach (1956), p. 33.
[10] Williams explicitly discusses this test in Williams (1972), pp. 41-42.
[11] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 7 ff.; Moore, *Principia Ethica*, rev. ed. (Cambridge: Cambridge University Press, 1993), pp. 59 ff.

> such thing as being just good or bad, there is only being a good or bad so-and-so.[12]

This seems a significant conclusion. Moore contended that we may sensibly ask of certain things (such as beauty, knowledge, pleasure) whether they are good and sensibly answer that indeed they are good.[13] But Geach claims that nothing can be just good in this way; what is good is always good *relative* to a certain kind. A good car may be a bad investment, a good cricketer may be a bad person.

Thomson has elaborated considerably on Geach's thesis. She endorses his claim that "good" is logically attributive, though not his conclusion that whatever is good is good relative to a certain kind. For although "That's good" may mean "That's a good *K*," for some kind *K*, it may not. When we say, for instance, that something is "good *to* eat (look at, listen to)," or "good *for* use in making cheesecake (hammering in nails, Alfred, Alfred's lawn mower)," or that someone is "good *at* hanging wallpaper (playing chess, singing)," or "good *in Hamlet* (the play)," or "good *as* Hamlet (the Prince)," or "good *with* children," and so on, we aren't necessarily implying that the thing or person in question is good relative to a certain kind. ("For what *K*," she asks, "could it at all plausibly be thought that being good for Alfred is being a good *K*?"[14]) Rather, all such uses of "good plus adjunct" involve what she calls *first-order ways of being good*,[15] and being good relative to a certain kind is itself reducible to being good in some first-order way. For example, "*x* is a good book" may mean that *x* is good to read, or good for use in teaching logic, or good to look at, etc.; which of these it means will depend on the context of its utterance.

Thomson also identifies what she calls *second-order ways of being good*.[16] If, for example, you say that a certain act was good, you might mean that it was good in some first-order way (e.g., good to do, or good to look at), but you're far more likely to mean that it was just, or generous, or kind, and so on. Though distinct from first-order ways of being good, these second-order ways of being good, Thomson contends, nonetheless "rest on" first-order ways of being good. Furthermore, she claims, there are no kinds of ways of being good other than being good in some first-order way or being good in some second-order way.[17]

Now, distinguishing these (kinds of) ways of being good is not itself anti-Moorean in spirit. But it may well seem so when coupled with the following crucial contention: if something is good, then it is good in some such way. Or, as Thomson puts it (I'll call this Thomson's First Thesis): "[A]ll goodness is goodness in a way."[18] Even though she disagrees with certain aspects of Geach's view, then, Thomson is nonetheless very much in sympathy with its spirit. She firmly embraces the idea that goodness is always

[12] Geach (1956), p. 34.
[13] This is only a very rough summary of his view. See Moore (1903) [(1993)], ch. 6. This view will be discussed more fully in later chapters.
[14] Thomson (1997), p. 278 [* p. 135 of this volume].
[15] Thomson (1997), pp. 276 ff. [* pp. 134 ff. of this volume].
[16] Thomson (1997), pp. 279 ff. [* pp. 136 ff. of this volume].
[17] Thomson (1994), p. 11.
[18] Thomson (1997), p. 276 [* p. 133 of this volume].

to be relativized in some way; it cannot stand alone. And this leads her to proclaim what I'll call Thomson's Second Thesis: "In saying 'That's good', we are not ascribing a property goodness – indeed, there is no such thing."[19]

2. AN UNWARRANTED GENERALIZATION?

We surely shouldn't doubt the plausibility of Geach's distinction between logically predicative and attributive adjectives (initial plausibility, at least; I'll express reservations about it in Section 4 below); and we shouldn't doubt the plausibility of Thomson's emphasis (reminiscent of Georg von Wright's[20]) on the variety of ways in which something can be good. Nonetheless we should resist the conclusion that there is no such thing as the sort of goodness with which Moore is concerned.

It's worth noting that the distinction that Geach invokes was discussed more than a quarter of a century earlier by W. D. Ross.[21] Ross endorsed this distinction but was nonetheless a firm supporter of the concept of intrinsic value. He recognized that "good" is frequently used in a logically attributive (or, as he sometimes put it, adjunctive) way, and he claimed that when so used it is relative in two respects: first, it's relative to the kind in question (for a good *x* may be a bad *y* or *z*, as we've already noted); secondly, it's relative to other members of the kind in question, in that for an x to be a good *x* it must (he contended) be better, perhaps considerably better, than most *x*s. (There's obviously some confusion here, for it is perfectly consistent to say, for instance, that most, or even all, knives are good knives.[22] Thus the second alleged type of relativization is clearly a fiction. But the first type equally clearly is not.) Ross also claimed, however, that "good" can be and sometimes is used in a logically predicative way, and that when so used it is not relative in either of these respects. When used as a logical predicative, he said, "good" expresses either intrinsic goodness (as in "Pleasure is good") or some related notion such as instrumental goodness (roughly, that type of goodness which consists in being conducive to something intrinsically good – see the appendix for a fuller discussion) or ultimate goodness (roughly, that type of goodness which consists in being intrinsically good while having no part that is not intrinsically good).

In light of Ross's discussion of the uses of "good," it is initially tempting to respond to the claims of Geach and Thomson as follows: "You're quite right to say that 'good' sometimes operates in the ways you mention. But so what? It would clearly be fallacious to infer from this that 'good' never operates in some other way. And in fact it does sometimes operate in the ways that Moore mentions. Thus your observations constitute no indictment of his conception of goodness."

---

[19] Thomson (1997), p. 275 [* p. 133 of this volume].
[20] See Georg Henrik von Wright, *The Varieties of Goodness* (London: Routledge and Kegan Paul, 1963), to which Thomson frequently expresses her indebtedness.
[21] See ch. 3 of W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930).
[22] See Thomson (1994), p. 12. Perhaps it is somewhat more plausible to contend that a good knife must be (considerably) better than the average possible knife, but I'm not at all sure how one is supposed to count such things.

But this response would be facile. After all, Geach (presumably[23]) and Thomson (certainly[24]) are quite aware of Ross's discussion but nonetheless reject the claim that "good" sometimes operates in the ways Moore mentions. Why?

3. GENERIC GOODNESS

The reason can, I think, be traced at least in part to a misleading practice of Moore's. Despite his laudable efforts at precision, Moore sometimes writes rather loosely, saying simply "good" when what he means (or should mean) in particular is "intrinsically good." (This is a very common practice.[25] I have in fact deliberately engaged in it so far in this chapter, since doing so helps make the case for the opponents of intrinsic value seem stronger than it really is.) At the very beginning of *Principia Ethica* Moore says that the question he will address is the question, "What is good?" (note: not the question, "What is *intrinsically* good?"), and he goes on to say: "'Books are good' would be an answer to it, though an answer obviously false; for some books are very bad indeed."[26] This puzzles Thomson, and understandably so. It doesn't look as if Moore is here concerned with intrinsic goodness, for who has ever been tempted to claim that books in general are intrinsically good? And, since no way of being good is specified, it may seem that Moore is concerned with some generic form of goodness. But this too is "weird" (as Thomson puts it[27]). Again, who has ever been tempted to claim – indeed, what could it mean to say – that books in general are "just plain good"?

There is no doubt that Thomson is in large part concerned to deny that there is any generic form of goodness – that there is a property, as she sometimes puts it, of "pure, unadulterated" goodness.[28] This, indeed, is what I have called her Second Thesis. Now, it may be that Moore thinks there is such a property. In fact, in *Ethics* he appears to offer a definition of intrinsic goodness in terms of generic goodness when he says:

> By saying that a thing is intrinsically good it [that is, the theory, utilitarianism, that Moore is investigating and which he endorses at least in this respect] means that it would be a good thing that the thing in question should exist, even if it existed *quite alone*, without any further accompaniments or effects whatever.[29]

But whether or not Moore would accept the existence of a property of generic goodness, we should note that it is perfectly consistent to deny its existence while

[23] See Geach (1956), p. 42, n. 1, where Geach refers to another chapter of Ross (1930).
[24] See Thomson (1994), p. 18, n. 4.
[25] Ross sometimes engages in it; see, e.g., Ross (1930), ch. 5 (but here he is careful to ensure that the reader doesn't forget that it is intrinsic goodness in particular with which he is concerned). Many others do too. See, e.g., Panayot Butchvarov, *Skepticism in Ethics* (Indianapolis: Indiana University Press, 1989), chs. 3 and 4.
[26] Moore (1903), p. 3 [(1993), p. 55].
[27] Thomson (1997), p. 276 [* p. 133 of this volume].
[28] Thomson (1996), pp. 129-30.
[29] G. E. Moore, *Ethics* (Oxford: Oxford University Press, 1912), p. 27. Compare Roderick M. Chisholm, "Objectives and Intrinsic Value," in *Jenseits von Sein und Nichtsein*, ed. R. Haller (Graz, Austria: Akademische Druck- und Verlagsanstalt, 1972), p. 262 [* p. 172 of this volume]. Moore would deny that such a "definition" constitutes an "analysis," since he declares the concept of intrinsic goodness unanalyzable. This is an issue I'll address in ch. 4.

nonetheless accepting the existence of *intrinsic* goodness. Moreover, it is perfectly clear that Moore's main concern is (no matter how he may at times express himself) with intrinsic goodness in particular rather than with goodness in general. Thus, even if Thomson's skepticism concerning the existence of "pure, unadulterated" goodness is on target, this would seem to leave the concept of goodness with which Moore is concerned wholly unscathed.

But we should ask: is there indeed no such thing as generic goodness? Thomson appears to believe that her First Thesis (that all goodness is goodness in a way) implies her Second Thesis (that there is no such thing as generic goodness). Does it?

Consider the analogy with shape. (Color would provide another analogy.) It's plausible to say that all shape is shape in a way, that nothing can have "pure, unadulterated" shape. What does this mean? It seems to mean (at least) that shape is a *determinable* property; nothing can have a shape without having a particular shape. Yet a determinable property is, of course, a property – a property, indeed, which unites otherwise dissimilar things. (Shape is something common to both squares and circles.) Why not take generic goodness to be on a par with shape in this regard?[30]

At one point Thomson entertains the idea that there is a property of goodness after all, namely, that long disjunctive property of being either good in this way or good in that way or...[31] Let's call this the property of being good in some way or other. In fact, Thomson says that there doubtless is such a property but then says that, since it's not what anybody ever means in saying "That's good," its title to being called "goodness" is dubious. This is odd. People don't typically say "That's shaped," but we can imagine this expression coming into vogue (among car enthusiasts, say, or body-building fans). If it did, it wouldn't be used simply to mean that the thing in question has some shape or other. But *this* hardly shows that the property of having some shape or other has no title to being called "shape."

Nonetheless, I think we must accept that the property of having a shape is not just the property of having some shape or other. After all, if the notion of property is construed liberally enough, there's presumably a property corresponding to the disjunction of any properties you like. There is, for example, the property of being either a number or a brick. What's strange about this property is that its disjuncts have nothing in common (or nothing interesting; I suppose they share the property of involving an entity, the property of being used in an example by me, and so on). The property of having some shape or other isn't like this, though, for its disjuncts do have

---

[30] The analogy may be flawed (as all analogies are in one way or another) in the following respect. It seems natural to say that something can have only one shape (or color) at once, whereas something might be good in several ways at once. But this is difficult. Something can have only one "overall" shape at once, but its parts can be variously shaped. Perhaps we can also say that something can have only one "overall" value at once, thereby preserving the analogy. (Overall value is discussed briefly in the appendix.) Also, abstracting from shape (and color), we can clearly say that something can have more than one visual quality at once, thereby perhaps preserving the analogy in a different way. But whether or not either of these moves is acceptable, it certainly seems reasonable to think that shape (or color) and goodness are analogous at least in respect of being determinable properties

[31] Thomson (1997), p. 277 [* p. 134 of this volume].

something interesting in common: they all involve shape. So shape comes first, as it were; having some shape or other comes after.

Something similar can be said about goodness. This is presumably distinct from being good in some way or other, since the latter presupposes the former. But then the former does exist, after all.[32] Indeed, Thomson seems to be uneasily aware of this herself. She says:

> [I]t is not mere happenstance that the word "good" appears in the expressions "good to look at", "good at hanging wallpaper",...and so on: its meaning makes a contribution to their meanings... Intuitively, for a thing *X* to be good in one of the first-order ways is for *X* to *benefit* someone or some thing *Y*...in the appropriate way, or to be capable of doing so.[33]

Of course, she believes that all benefit is benefit in a way; but that doesn't alter the fact that all these ways have something in common, namely, they are all ways of *being a benefit*. Yet on the very same page Thomson once again declares: "[T]here is no such property as goodness." These statements are difficult to reconcile with one another.

But all this is strictly by the by. For, regardless of whether we agree with Thomson's Second Thesis concerning the existence of some form of generic goodness, the essential point remains: whether or not there is such a property as *intrinsic* goodness is an independent issue.

4. GEACH'S TESTS REVISITED

Keeping in mind this emphasis on intrinsic goodness in particular, let's return for a moment to Geach's twin tests. Thomson's First Thesis is that all goodness is goodness in a way. I have not disputed this. In light of this, it's tempting to diagnose the reason for the failure of "good" to pass Geach's tests as follows: no particular way of being good has been specified. Once a way of being good has been specified, passing the tests seems plain sailing. Consider: ***x* is an apple that is good to eat** would seem to split up into ***x* is an apple** and ***x* is good to eat**. Also, ***x* is an apple that is good to eat** and **an apple is a fruit** would seem jointly to imply ***x* is a fruit that is good to eat**. Similarly, ***x* is a painting that is good to look at** would seem to split up into ***x* is a painting** and ***x* is good to look at**. Also, ***x* is a painting that is good to look at** and **a painting is a work of art** would seem jointly to imply ***x* is a work of art that is good to look at**. Now consider the following: ***x* (e.g., conscientiousness) is an intrinsically good state of mind**. Doesn't it seem correct to say that this indeed splits up into ***x* is a state of mind** and ***x* is intrinsically good**? Likewise, doesn't it seem correct to say that ***x* is an intrinsically good state of mind** and **a state of mind is a state of affairs** jointly imply ***x* is an intrinsically good state of affairs**? Isn't this evidence that being intrinsically good is just a way of being good, as are being good to eat and being good to look at, constituting at

---

[32] This is not to say that the sort of definition of intrinsic goodness in terms of generic goodness apparently proposed by Moore in the passage in *Ethics*, quoted above, is acceptable. I think it is unacceptable. I'll return to this issue in chs. 4 and 5.

[33] Thomson (1997), p. 289 [* p. 144 of this volume].

once a confirmation *both* of Thomson's First Thesis *and* of the existence of intrinsic value?

Unfortunately, matters aren't quite that simple. First, there's reason to feel uneasy about the claim that being intrinsically good is a way of being good, as being good to eat and being good to look at are (and hence reason to feel uneasy about the relations just alleged between ***x* is an intrinsically good state of mind** and the other phrase-forms mentioned). I'll return to this in the next section. Of more immediate concern, though, is that there is reason to doubt the relevance of Geach's tests to the issue at hand, intriguing as they are. Thomson herself expresses such doubts. Concerning Geach's claim that, whereas "good" is logically attributive, "red" is logically predicative, she says:

> "Red" was not in fact well chosen for Geach's purposes, since "red" is heavily context dependent: what we ascribe to an apple when we say "It's red" is different from what we ascribe to the paint in a certain can when we say "It's red". Better choices would have been "visible" or "happy"…or..."sauteed"...or "poisonous"…or...[34]

Now, it seems quite right to say that what's red as far as apples go may not be red as far as paints go. But of course this sort of division can be continued. After all, what's red as far as Macintosh apples go may not be red as far as Red Delicious apples go. And, contrary to what Thomson seems to imply, what's visible or poisonous to *x* may *not* be visible or poisonous to *y*. The fact is, *very many* properties are determinable (to some extent) rather than (fully) determinate, including all those just mentioned; and, as far as I can tell, every less-than-fully determinate property *A* is such that we can imagine both possible circumstances in which we would balk at inferring ***x* is a *B*** and ***x* is *A*** from ***x* is an *A B*** (although imagining such cases may be harder the more determinate the property is) and also possible circumstances in which we would not balk at this (although imagining such cases may be harder the less determinate the property is). Whether or not we do in fact balk at such inferences would seem to depend, then, on contingent circumstances (e.g., that the different ways of being red happen not to be – or, at least, happen not to be taken to be – importantly varied) which have nothing to do with the strict logical propriety of such inferences. This suggests that Geach's tests are simply irrelevant, pointing up no essential differences between the properties expressed by "red" and "good" and revealing no important insights into the nature of these properties.[35] (It should be noted that, while Thomson endorses Geach's claim that "good" is logically attributive and elaborates on this, she doesn't dwell on these tests.)

5. KINDS OF WAYS OF BEING GOOD

Even if the relevance of Geach's tests is suspect, though, they do suggest that being intrinsically good is just one way of being good. Why not, then, declare the existence of

---

[34] Thomson (1997), p. 277, n. 5 [* pp. 134-35 of this volume].

[35] The tests, we might say, have been a red herring – though neither red nor herring.

intrinsic value to be perfectly compatible not only, as I argued in Section 3 above, with Thomson's Second Thesis (that there is no such thing as generic goodness) but also with her First Thesis (that all goodness is goodness in a way)?

I take it that Thomson would resist this. I think that she'd say that she has in mind a *particular kind* of way of being good, and that something's being intrinsically good (were this possible) would not be a case of its being good in this particular kind of way. We should now look into this.

Thomson is perfectly happy to acknowledge a certain distinction on which proponents of intrinsic value rely, the distinction between (as she puts it) nonderivative and derivative goodness. This distinction, she says, cuts across ways of being good.[36] Something is derivatively good in some way if, as she puts it, it "inherits" its goodness from something else; otherwise, its goodness is nonderivative. (This analogy of inheritance would in fact seem less appropriate in some cases than in others. See the appendix.) It may be, for instance, that drinking lemonade is good for Alfred because it quenches his thirst and his thirst's being quenched is good for him, in which case his drinking lemonade is derivatively good for him. It may also be, of course, that his thirst's being quenched is good for him only derivatively. (Whether this is so will depend on just what nonderivative goodness-for-persons consists in, a difficult question that we needn't address here.) Likewise, it may be that lubrication is good for Alfred's lawn mower; if so, it would seem that it is so only derivatively. Of course, lubrication wouldn't be good for Alfred, and the application of lemonade wouldn't be good for his lawn mower; goodness-for-persons consists in something different from what goodness-for-artifacts consists in. This leads Thomson to an important observation: since the distinction between nonderivative and derivative goodness cuts across ways of being good, it cannot itself denote ways of being good (in the sense of "way" at issue). As she puts it:

> [D]erivative goodness is not itself a way of being good. A derivatively good thing is a thing that inherits goodness from something, and what it inherits is not derivative goodness, but rather goodness in way W, for the appropriate W. What it inherits might be goodness for this or that person, or [some other way of being good]... Since derivative goodness is not a way of being good, neither is nonderivative goodness.[37]

I believe that it's absolutely correct both to distinguish between nonderivative and derivative goodness and to deny that they constitute ways of being good, in the sense of "way" at issue. (Thus, if "nonderivatively" were substituted for "intrinsically" in ***x* is an intrinsically good state of mind**, we should clearly *deny* that this sanctions splitting up and transposition in the manner discussed in the last section.) It's also correct to observe that proponents of intrinsic value rely on this distinction when contrasting intrinsic value with extrinsic value. But of course this would show that being intrinsically good is not (or, better, does not involve) a way of being good, in the sense of "way" at issue, only if proponents of intrinsic value claimed that being intrinsically good just *is* being

[36] Thomson (1992), pp. 99 ff.
[37] Thomson (1992), p. 103.

nonderivatively good; and it's clear that, although someone is of course free to use the term "intrinsically good" to mean simply what Thomson means by "nonderivatively good," this is *not* how the proponents of intrinsic value typically use the term (as Thomson herself acknowledges[38]). Rather, they take being intrinsically good to be a particular way in which something can be nonderivatively good; being extrinsically good involves being derivatively good *in that same way*. (Given this, the claim that ***x* is an intrinsically good state of mind** sanctions splitting up and transposition does indeed seem no more problematic than the other cases mentioned in the last section.)

The key question then is: *what* way of being good does being intrinsically good involve? Here we'll probably meet with some disagreement. It's certainly true that proponents of intrinsic value haven't traditionally concerned themselves with specifying in any great detail the way of being good that being intrinsically good involves. Indeed, they may seem at times positively hostile to the suggestion that there is some such way. (Recall Ross's insistence on the nonrelativity of "good" when used in a logically predicative way. Consider also Moore's use of the term "good absolutely" to express the idea of intrinsic goodness.[39]) Yet it's perfectly clear that the proponents of intrinsic value have often been at pains to distinguish the various senses of "good" and to insist that "intrinsically good" expresses only one of these senses.[40] And if

---

[38] Thomson (1992), pp. 106-7. For an example where the term does appear to be used in this way, see L. W. Sumner, *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press, 1996), p. 48.

[39] Moore (1903), p. 99 [(1993), p. 150]. This occurs in the context of Moore's famous argument against the coherence of ethical egoism. There he insists that intrinsic value is not "private" or person-relative. This is surely correct, given that the intrinsic value of something supervenes solely on its intrinsic nature (a matter to be explored in the next chapter). Such *non-relative* goodness may reasonably be called *absolute* goodness, but that should not lead us to think, as Moore appears at least at times to have thought, that there aren't different ways in which something can be good. Clearly there are. Moore's argument against ethical egoism overlooks this fact. That intrinsic value isn't person-relative is no reason to think that there are no kinds of value that are. Even if it makes no sense to say that something is *intrinsically* good relative to Alfred but not to Bert, it's perfectly possible that it should be good in some *other* way relative to Alfred but not to Bert. If ethical egoism is understood in terms of some such person-relative value rather than in terms of intrinsic value, Moore's argument against it simply misses the mark. *Is* there some such value? It's surely reasonable to think so. It certainly seems that what is in one person's *interests* needn't be in another's, and if, as seems reasonable, we say that what is in a person's interests is "good *for him*," then we're invoking a kind of value that is explicitly person-relative. We can, as is customary, call this kind of value "prudential value" to distinguish it from intrinsic value. (Perhaps it sometimes happens that someone uses the term "intrinsically good" to express nonderivative prudential goodness rather than what I have called intrinsic goodness. If so, I wouldn't accuse such a person of talking nonsense – as Moore apparently would – but I would accuse him of misusing the term in question.) See Sumner (1996), pp. 46 ff., for a criticism of Moore's argument against ethical egoism that accuses him of failing to distinguish between prudential and intrinsic value.

[40] This is true, I think, even though the term "intrinsic value" is sometimes used to refer to more than one kind of value. On p. 260 of G. E. Moore, *Philosophical Studies* (London: Routledge and Kegan Paul, 1922), Moore says that to say that a kind of value is intrinsic is to say that its possession depends solely on the intrinsic nature of the thing that possesses it, and he goes on to declare that both goodness and beauty are kinds of intrinsic value. But notice that even then he restricts the term "intrinsically good" to just one kind of intrinsic value. Unlike Moore here, but like many others, in this book I will restrict my use of "intrinsic value" to refer to intrinsic goodness, neutrality, and badness; even if beauty depends solely on the intrinsic nature of that which possesses it (something that seems to me quite dubious), I will not be referring to it when I talk of intrinsic value. (Nonetheless, it may of course be that beauty is itself intrinsically good.) These issues will receive fuller treatment in the next chapter.

"intrinsically good" doesn't reduce to "nonderivatively good," it would seem legitimate to attribute to the proponents of intrinsic value the view that being intrinsically good involves just one way of being good (in Thomson's sense of "way"). But again, what way is this?

The answer is: *ethical goodness*. When it's said that beauty, or knowledge, or pleasure, or virtue is *intrinsically* good and that, for example, activities that promote such states are *extrinsically* good, what's meant is that all these things are *ethically* good. Now, I hasten to add that not all of these goods are *moral* goods, as this term is usually understood. Virtue is plausibly thought of as a moral good, but beauty, knowledge, and pleasure presumably are not.[41] What then can it mean to say that these nonmoral goods are nonetheless ethically good? This division of terms is admittedly awkward, since I'm not trying here to draw any distinction between ethics and morality. But the point is that even those intrinsic goods such as beauty, knowledge, and pleasure (assuming they are such), which are not normally thought of as moral goods, nonetheless have an intimate tie to morality, in that *there is a moral requirement to favor them* (welcome them, admire them, take satisfaction in them, and so on) *for their own sakes*. That which is intrinsically good is preferable to that which is not, the "-able" here expressing *moral* worthiness. (This is an issue that I'll discuss in detail in Chapter 4.)

At this point I need to pause and take stock of certain distinctions. I have talked of "intrinsic goodness" and of "nonderivative ethical goodness," and just now I alluded to something's being "ethically good for its own sake." Just how are these related?

First, it would be a mistake to say that what is intrinsically good is nonderivatively ethically good, although quite correct to say that what is extrinsically good is derivatively ethically good. This is because only that which is *basically* intrinsically good is nonderivatively ethically good; that which is *non*basically intrinsically good is derivatively good.[42] (I shall address the distinction between basic and nonbasic intrinsic value in Chapter 5, when discussing the computation of intrinsic value. There it will be maintained that the intrinsic value of a "whole" is to be computed by summing the intrinsic values of those "parts" of it that have basic intrinsic value. The value of the whole is thus derivative from the values of its parts. The manner in which extrinsic value is derivative from intrinsic value is different, as I explain in the appendix.)

Second, ever since Christine Korsgaard published her influential paper on two distinctions in goodness, it has been recognized that we must distinguish between the value that something has in virtue of its intrinsic nature and the value that it has for its own sake.[43] The former she calls "intrinsic value," the latter "final value." I am happy to concede that my primary concern in this book is to give an account of the nature of

---

[41] Regarding pleasure, this is in fact an oversimplification. See ch. 6.

[42] Cf. Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "A Distinction in Value: Intrinsic and For Its Own Sake," *Proceedings of the Aristotelian Society* 100 (1999), n. 14 [* p. 120 of this volume], where credit for this observation is given to Bengt Brülde.

[43] Christine M. Korsgaard, "Two Distinctions in Goodness," *Philosophical Review* 92 (1983): 169-95 [* pp. 77-96 of this volume], and Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), ch. 9.

final ethical value, that is, of what it is for something to be ethically valuable *in its own right* or *for its own sake*. But, as I will argue in Chapter 3, whatever has such value *does* have it in virtue of its intrinsic nature. And so, when it comes to ethical value at least, there is no reason not to continue to use the more traditional term "intrinsic value" where Korsgaard would talk of "final value."[44]

In saying that intrinsic value and extrinsic value are species of *ethical* value, I don't mean to say that all proponents of intrinsic value will readily agree with this claim. Some, indeed, may be inclined to reject it; if so, I can only hope that further reflection will show them the error of their ways. Be that as it may, the point is certainly not original with me. Many philosophers have made it.[45] The title of Moore's masterpiece is, after all, "Principia *Ethica*." And, despite his insistence that "good," when used in a logically predicative way, is not relative in the ways in which it is relative when used in a logically attributive way, Ross understands the point perfectly well. He puts it nicely in the opening paragraph to his chapter on moral goodness in *Foundations of Ethics*:

> I have suggested that things that are good in the predicative as opposed to the adjunctive sense fall into two classes: (1) those that are good in the sense of being worthy objects of admiration, and (2) those that are good in the sense of being worthy objects of satisfaction. Both of these come, from one point of view, within the scope of ethics; for a thing's being good in either of these ways brings into being a *prima facie* [moral] obligation to produce that thing... But goods of the second type are not themselves, as such, morally good. Nor, again, are all goods of the *first* type themselves as such morally good; excellent scientific or artistic activity is good but not morally good.[46]

It's an interesting question, of course, what distinguishes those intrinsic goods that are moral goods from those that are not. We needn't address this question here, however (I'll return to it in Chapter 6). The important point for now is simply that what is

[44] Moore himself might balk at this. In Moore (1912), pp. 30-31, he says that "good for its own sake" is necessarily coextensive with "ultimately good" but not with "intrinsically good." (Contrast, however, Moore's *Philosophical Papers* (London: Allen and Unwin, 1959), p. 94: "I use the phrase 'intrinsically good' to mean precisely the same as 'worth having for its own sake'…") That which is intrinsically good might have a part that is not intrinsically good; that which is ultimately good can have no such part. But, whereas this distinction can of course be drawn, I'm not sure why we should think it an important one (except in the particular case where what is ultimately good is so because it is basically intrinsically good – see ch. 5). More to the point, I see no reason to think that it is a mistake to say that what is intrinsically good but not ultimately good is good for its own sake.

[45] See, among others: Franz Brentano, *The Origin of Our Knowledge of Right and Wrong* (London: Routledge and Kegan Paul, 1969); Moore, Ross, Ewing, and Chisholm, in several works, including those cited above; Noah Lemos, *Intrinsic Value* (Cambridge: Cambridge University Press, 1994); and Robert Audi, "Intrinsic Value and Moral Obligation, *Southern Journal of Philosophy* 35 (1997): 135-54. Also, although Sumner uses the term "intrinsic value" liberally (see n. 38 above), he explicitly notes that Moore's use of it in *Principia Ethica* is restricted to the ethical dimension of value.

[46] W. D. Ross, *Foundations of Ethics* (Oxford: Oxford University Press, 1939), p. 290. It may be that the penchant of some writers, such as Moore, to use "good" without qualification or "good absolutely" to express intrinsic goodness in particular is to be explained by their holding the view (I don't know whether this is true of Moore) that ethical or moral values somehow take precedence over other values. Far from this implying that being intrinsically good is not a way of being good, it presupposes that it is – and that it is a dominant way.

intrinsically good is good in a certain way – it is what I have called ethically good; there is an ethical requirement that one favor it (for its own sake) – and thus that accepting Thomson's First Thesis provides no reason to reject the concept of intrinsic value.

It's interesting to note that this account of intrinsic value would appear to be perfectly in keeping with the observations of Philippa Foot, who expresses some doubts about the concept. Foot is a critic of consequentialists (such as Moore), claiming that "we go wrong in accepting the idea that there *are* better or worse states of affairs in the sense that consequentialism requires."[47] She says this because she thinks that consequentialism requires a sense of "good" that is divorced from morality, in terms of which "morally right" may be given an account, and no such sense can be found (although she believes that "good" can be given a sense "*within* morality," as she puts it).[48] But this misconstrues consequentialism, I think (a misconstrual of which consequentialists themselves may sometimes have been guilty). For although it's accurate to say that consequentialists seek to account for "morally right" (in the sense that concerns them: *overall*, rather than merely *prima facie*, moral rightness) in terms of "good," it's nonetheless the case (I contend) that the sense of "good" at issue is what I've called ethical goodness. Foot herself would in fact seem perfectly receptive to this idea. While finding fault with both consequentialism and the idea that some states of affairs are good or bad "from a moral point of view," she nonetheless confesses that it "seems preposterous…to deny that there are some things that a moral person must want and aim at in so far as he is a moral person and that he will count it 'a good thing' when these things happen…"[49] But this is all that's at issue when something is intrinsically good.

Whether or not Foot would in the end agree with me, however, Thomson would presumably continue to disagree. At one point she entertains the idea that moral goodness is a way of being good, and that that is the way (or one way) in which compassion is good.[50] But she drifts away from this idea, neither pursuing it nor decisively refuting it. She appears to doubt the possibility of anyone's giving a clear account of what moral goodness consists in, and then moves on. But this hardly settles the matter. Whether she has in mind (as I think she does) "morally good" in the sense in which virtue is often said to be morally good but beauty not, or the broader sense that I have invoked in which both virtue and beauty may (with some plausibility, at least) be said to be ethically good, Thomson's inability to come up with a clear account of what such goodness consists in of course does not imply that there is no such account to be given. And even if there is no such account to be given, this hardly shows that the goodness in question doesn't exist. It may be that the reason is, rather, that the type of goodness in question is unanalyzable,[51] so that to the question, "What is it that moral

---

[47] Foot (1985), p. 199.
[48] Foot (1985), pp. 205-6.
[49] Foot (1985), p. 204.
[50] Thomson (1992), p. 102.
[51] I will in fact dispute this contention in ch. 4 (with respect to ethical goodness in the broader sense) and in ch. 6 (with respect to moral goodness in the narrower sense).

goods have in common?" one can give only the unenlightening answer, "They're all morally good." (Compare the question, "What is it that red things have in common?")

Thomson might still demur. Recall her contention that, for something to be good in one of the first-order ways, it must *benefit* someone or something, or be capable of doing so.[52] Goodness, she is thereby claiming, is always relative to some category of object. But to what category of object is the alleged way of being good that I've called ethical goodness relative? If there is none, then there is no such way of being good after all.

I suppose that some might seek to respond to this by proposing some category of object to which ethical goodness is relative. The obvious category here would be persons. (Of course, somehow the distinction between being intrinsically good and being good-for-persons would have to be preserved, so that what is intrinsically good is good relative to persons in a way different from that in which Alfred's well-being is good-for-Alfred.) Indeed, person-oriented accounts of morality are extremely common (social contract theories being a prominent example[53]). If some such account of ethical goodness is correct, then clearly the existence of intrinsic value is consistent with its being true that all first-order ways of being good involve benefit. However, I find Thomson's contention dubious, and, in the absence of any argument for it (Thomson simply relies on her intuition[54]), there's no need to endorse any such account here. But what, then, might be said to unite the ways of *being good*, making them ways of being good, if it's not the idea of benefit? My answer is: the idea of *valuableness* (that is, the worthiness of being valued). I grant that this answer may not at first seem particularly illuminating, but in Chapter 4 I'll try to show that it is actually quite helpful at least with respect to clarifying the concept of *intrinsic* goodness.

6. ORDERS OF WAYS OF BEING GOOD

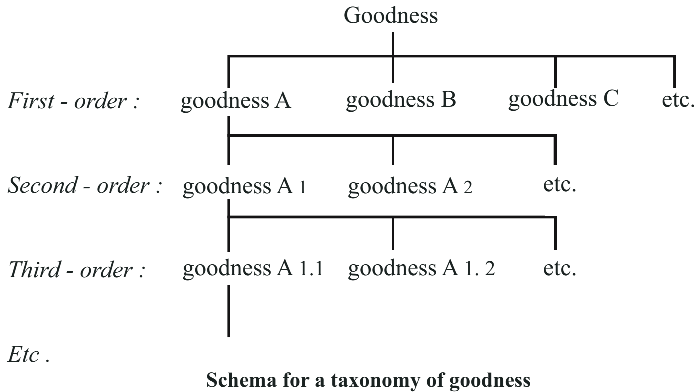
What of the *orders* of ways of being good? Thomson distinguishes two. I'm not sure just what should be said here. The intuitive picture is given in the following chart:

---

[52] This contention is endorsed by John Kekes on p. 61 of "On There Being Some Limits to Morality," in *The Good Life and the Human Good*, ed. E. F. Paul et al. (Cambridge: Cambridge University Press, 1992), pp. 63-80.

[53] Another example is what in Sumner (1996) is called "welfarism," according to which (roughly) intrinsic value is grounded in prudential value in the sense that nothing is good in an ethically relevant sense unless it is good for someone. Sumner defends this claim in the final chapter of his book.

[54] Thomson (1997), p. 289 [* p. 144 of this volume].



Schema for a taxonomy of goodness

But where in this picture we should place a particular way of being good is not clear; some serious theorizing is called for. My inclination is to call ethical (both intrinsic and extrinsic) goodness a first-order way of being good, because I don't know what could be said to mediate between it and generic goodness. I'm inclined to call moral goodness (of the sort exemplified by virtue but not by beauty) a second-order way of being good (for it is to be subsumed under ethical goodness), and certain particular goods (such as those ways of being good that Thomson calls second-order: being just, generous, kind, and so on) third-order ways of being good (for they are to be subsumed, I believe, under moral goodness. Thomson's account of what makes these ways of being good non-first-order ways of being good is of course different[55]). But I won't try to argue here for what I'm inclined to say in this regard; for, while there are obvious problems to be resolved (Might not an A-type goodness also be subsumable under B? Might it not even be subsumable at a different level under B than under A? Etc.), my contention that being intrinsically good involves a way of being good can be accepted and investigated even in the absence of an account of what order of way of being good this is.

But *should* this contention be accepted? I acknowledge that I haven't argued for it here. Indeed, I don't know how to do so. I have argued that certain critics of the concept of intrinsic value have been unsuccessful, but that of course is not tantamount to a positive argument in favor of the concept. I can only resort to the following plea: isn't it clear, on reflection, that a morally sensitive person will favor certain things, but not others, for their own sakes? If so, that's basically all you need to accept that some things have intrinsic value. But to give you a further nudge toward acceptance of this, let me end this chapter by asking you to reflect on a pithy comment by Panayot Butchvarov, who says, when discussing Geach's claim that "*x* is good" requires completion by means of a statement of the form "*x* is a good *F*":

[55] Thomson (1997), pp. 281-6 [* pp. 138-42 of this volume].

Nevertheless, millions have thought they understood Genesis 1:31: "And God saw every thing he had made, and behold, it was very good."[56]

[56] Butchvarov (1989), p. 17. I would ask you also to consider the following observation, blatantly *ad feminam* though it may be. At one point Thomson asks: "Is it plausible to think that what has intrinsic goodness just is what a person (all people?) would value for its own sake if he or she were fully informed, free of neuroses, and assessing the matter in a cool hour?" To which she responds: "No, unless we can show that people really would not love the nasty under this constraint." (Thomson (1992), p. 108.) This response appears to betray the fact that Thomson understands perfectly well what the proponents of intrinsic value take intrinsic value to be, and that she herself believes that some things have such value. For what else is "the nasty" supposed to refer to here, if not to that which is intrinsically bad?

PART III

IDENTIFYING THE CONCEPT OF INTRINSIC VALUE

CHAPTER 13

R. M. CHISHOLM

OBJECTIVES AND INTRINSIC VALUE

1.

Meinong made lasting contributions to almost every branch of philosophy. Among the most substantial of these are his writings on the theory of value, beginning with the *Psychologisch-ethische Untersuchungen zur Werttheorie*, in 1894, and ending with the posthumous *Zur Grundlegung der allgemeinen Werttheorie*, published in 1923, and the *Ethische Bausteine,* not published until 1968. Now, happily, Volume III of the *Meinong Gesamt Ausgabe* has appeared – *Abhandlungen zur Werttheorie*, published in Graz by the Akademische Druck- und Verlagsanstalt in 1968, and edited by the late Hofrat Dr. Rudolf Kindinger. (I regret very much that the Hofrat cannot be with us. All friends of Meinong and of Austrian philosophy are deeply indebted to him.) The *Abhandlungen zur Werttheorie* contain not only all of Meinong's published writings on the theory of value, but also the previously unpublished *Ethische Bausteine*, as well as extensive notes, addenda, and corrigenda that Meinong himself had made.

I will not attempt to summarize Meinong's contributions to the field. If one wants a general view of Meinong's place in the history of the theory of value, I would recommend these three writings: J. N. Findlay's, *Meinong's Theory of Objects and Values*, published by Oxford in 1963; Karl Wolf's "Die Grazer Schule: Gegenstandstheorie und Wertlehre" published in *Philosophie in Österreich*, edited by Leo Gabriel and Johann Mader (Vienna: Österreichischer Bundesverlag für Unterricht, Wissenschaft und Kunst, 1968); and Oskar Kraus's somewhat less sympathetic account in *Die Werttheorien: Geschichte und Kritik* (Brunn: Rudolf M. Rohrer, 1937).[1] What I will do – I hope in the spirit of Meinong, or better, in the spirit of Brentano, Meinong and Mally – is discuss some of the consequences of Meinong's theory of value. More exactly, I will discuss some of the consequences that follow if one makes certain general assumptions that Meinong had made in his

[1] See also Wolf's "Die Entwicklung der Wertphilosophie in der Schule Meinongs," in *Meinong-Gedenkschrift* (Graz: "Styria" Steirische Verlagsanstalt, 1952), ed., K. Radakovic, A. Silva-Tarouca, and F. Weinhandl, pp. 157-171; and Kraus's "Die Grundlagen der Werttheorie," in *Philosophische Jahrbücher*, II.Jg. (1914), pp. 1-48.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 171-179.
© 2005 *Springer. Printed in the Netherlands.*

later statements of his theory. I would say that the general assumptions are true, but that some of their consequences are rather different from what Meinong took them to be.

The general assumptions are these. (1) There is – in Ernst Mally's terms – a logic of the emotions, or of the will, which can be formulated with precision.[2] (2) It is only after formulating this logic that one can deal adequately with the philosophical problems that cluster around the concept of value. (3) The primary sense of the term "value" pertains to intrinsic value – to what Meinong called "impersonal value [*unpersönliche Wert*]"; this sense of the term is presupposed by the concepts of utility and instrumental value.[3] And (4) the bearers of intrinsic value – the objects to which value, in the primary sense of the term, may be ascribed – are all propositional objects. They are those objects, designatable by that-clauses, that were called "*Objektive*" by Meinong, "*Sachverhalte*" or "states of affairs" by Mally, and "propositions" by Russell.[4]

2.

What are the things that are of intrinsic value? Meinong speaks of "logical", "aesthetic", and "timological" value.[5] He has in mind the value of knowledge and insight, the value of beauty and fittingness, and, as I interpret him, the value that is to be found in pleasure and in the exercise of virtue. The contraries of these things are the things that are intrinsically evil: error and ignorance, ugliness and unfittingness, displeasure and pain and the exercise of vice. Here Meinong carries on the tradition of Brentano.

It is obvious that the terms we have just used – "knowledge," "beauty," fittingness," "pleasure," "the exercise of virtue," "error," "ignorance," "ugliness," "displeasure," "pain," "the exercise of vice" – do not designate *concreta.* In saying, for example, that *knowledge* is intrinsically good we mean, more exactly, that that state of affairs which is someone knowing something is intrinsically good. And in saying that pain is intrinsically bad, we mean that that state of affairs which is someone being in pain ("someone experiencing painfully") is one that is intrinsically bad.

And what do we mean when we say that a state of affairs is *intrinsically* good, or *intrinsically* bad – as distinguished from being merely *instrumentally* good or *instrumentally* bad? I suggest this: a state of affairs is *intrinsically* good if it is necessarily good – if it is good in every possible world in which it occurs. And a

---

[2] See Ernst Mally, *Grundgesetze des Sollens: Elemente der Logik des Willens* (Graz: Leuschner & Lubensky, 1926), esp. pp. 1-8.
[3] See *Abhandlungen zur Werttheorie*, pp. 280-282, 349, 404, 425 ff., 503, 625 ff.
[4] See Chapter III ("Das Objektiv") of Meinong's *Über Annahmen*, Second Edition, (Leipzig: Johann Ambrosius Barth, 1910), and Mally, loc. cit.
[5] See *Abhandlungen zur Werttheorie*, p. 639; also page 124 of Meinong's "Selbstdarstellung," in Raymund Schmidt, ed., *Die Deutsche Philosophie der Gegenwart in Selbstdarstellungen*, Vol. I (Leipzig: Felix Meiner, 1921), pp. 91-150.

state of affairs is *intrinsically* bad if it is necessarily bad – if it is bad in every possible world in which it occurs. Thus what is merely instrumentally good, good as a means, is something that happens to lead to good results in this world but does not lead to good results in every possible world.

Meinong held the following doctrine about the relation of value feelings (*Wertgefühle*) to intrinsic value. A state of affairs is intrinsically good if and only if *Seinsfreude* and *Nichtseinsleid* are the feelings that would be appropriate to it.[6] In other words, a state of affairs is intrinsically good if and only if joy is the feeling that would be appropriate to its being actual and sorrow is the feeling that would be appropriate to its not being actual. This doctrine has the following consequence: a state of affairs is intrinsically good if and only if its negation is intrinsically bad, and a state of affairs is intrinsically bad if and only if its negation is intrinsically good. And this consequence seems to me to be false.

Let us first see why the consequence follows. Let p be a state of affairs that is intrinsically good – say, Mr. Smith knowing something. Joy, then, will be the emotion that is appropriate to p's being actual and sorrow will be the emotion that is appropriate to p's not being actual. But, as Meinong himself emphasizes, p's being actual is equivalent to not-p's not being actual, and p's not being actual is equivalent to not-p's being actual. Or, more barbarically, the actuality of p is equivalent to the nonactuality of not-p, and the nonactuality of p is equivalent to the actuality of not-p.[7] Hence if sorrow is appropriate to the nonactuality of p, then sorrow is also appropriate to the actuality of not-p. But, according to Meinong's doctrine, if sorrow is appropriate to the actuality of a state of affairs, then that state of affairs is intrinsically bad. Hence the actuality of not-p is intrinsically bad. But the actuality of not-p is equivalent to not-p. And therefore, according to Meinong's doctrine, if a state of affairs is intrinsically good then its negation is intrinsically bad. And analogously: if a state of affairs is intrinsically bad, then its negation is intrinsically good.

If this doctrine were true, then every possible world would contain an infinite amount of good and also an infinite amount of evil.[8] Consider this world, for example. There are no men who are ten feet tall. Hence our world contains such goods as these: no ten-foot man being in pain, no ten-foot man being in error, no ten-foot man behaving wickedly, and similarly for eleven-foot men, twelve-foot men, and so on *ad indefinitum.* And our world contains such evils as these: no ten-foot man being in pleasure, no ten-foot man knowing anything, no ten-foot man behaving virtuously, and similarly for eleven-foot men, twelve-foot men, and so on *ad indefinitum.* But, surely, in adding up the goods in this world it is not appropriate to list such facts as there being no ten-foot man in pain, and in adding up the evils it is not appropriate to list such facts as there being no ten-foot man who knows

---

[6] See *Abhandlungen zur Werttheorie*, pp. 410 ff., 562 ff., 640.

[7] See Meinong's discussion of the "Koinzidenz der Untatsächlichkeit mit der Tatsächlichkeit des Gegenteiles" in *Möglichkeit und Wahrscheinlichkeit* (Leipzig: Johann Ambrosius Barth, 1915), p. 94 ff.

[8] Compare Kraus's criticism in *Die Werttheorien*, p. 227.

anything. A state of affairs is intrinsically good only to the extent that it entails the existence of pleasure, or knowledge, or beauty, or the exercise of virtue. But the nonexistence of a ten-foot man in pain does not entail any of these things and therefore it is not intrinsically good. And a state of affairs is intrinsically bad only to the extent that it entails the existence of displeasure, or error, or ugliness, or the exercise of vice. But the nonexistence of a ten-foot man who knows anything does not entail any of these things and therefore it is not intrinsically bad. Thus Aristotle said: "Positive goodness and badness are more important than the mere absence of goodness and badness: for positive goodness and badness are ends, which the mere absence of them cannot be."[9]

These facts have certain important consequences for the logic of the emotions. The negation of a good state of affairs will be neither good nor bad; and the negation of a bad state of affairs will be neither good nor bad. If by an *indifferent* state of affairs we mean a state of affairs having the same value as its negation, then the negations of good states of affairs and the negations of bad states of affairs will be themselves neither good, bad, nor indifferent. How, then, are we to relate the fundamental concepts of the theory of value?

One possibility is this. We take as our undefined concept that of one state of affairs being intrinsically preferable to another state of affairs. Then we set forth the following definitions. A state of affairs p is the *same in value* as a state of affairs q provided only p is not preferable to q and q is not preferable to p. A state of affairs is *indifferent* provided only it is the same in value as its negation. A state of affairs is *good* if it is preferable to a state of affairs that is indifferent, a state of affairs is *bad* if a state of affairs that is indifferent is preferable to it, and a state of affairs is *neutral* if it is the same in value as a state of affairs that is indifferent.

The following, I suggest, are plausible axioms. (Al) the relation of preferability is asymmetrical; (A2) the relation of not-being-preferable-to is transitive; (A3) all indifferent states of affairs are the same in value; (A4) all good states of affairs are preferable to their negations; (A5) all bad states of affairs have negations that are preferable to them; (A6) if a disjunction of two states of affairs is preferable to a given state of affairs, then either the one disjunct is preferable to the given state of affairs or the other disjunct is preferable to the given state of affairs; (A7) if a given state of affairs is preferable to a disjunction of two states of affairs, then the given state of affairs is preferable either to the one disjunct or to the other disjunct; (A8) every state of affairs that is entailed by the tautology, p or not-p, is indifferent; and finally, (A9) any two states of affairs that entail each other are the same in value.[10]

Among the consequences of these axioms are the following: sameness of intrinsic value is transitive, reflexive, and symmetrical; intrinsic preferability is transitive, irreflexive and asymmetrical; goodness, badness, and neutrality are

---

[9] *Rhetoric*, Book I, Ch. 7, 1364 a.

[10] The definitions of the value concepts above and the first five axioms are from Roderick M. Chisholm and Ernest Sosa, "On the Logic of 'Intrinsically Better'," *American Philosophical Quarterly*, Vol. III (1966), pp. 244-249.

exclusive and exhaustive; what is good is preferable to what is neutral; what is neutral is preferable to what is bad; whatever is preferable to what is good is itself good; whatever is bad is preferable to is itself bad; nothing that is good has a negation that is good; and nothing that is bad has a negation that is bad.

The neutral turns out to be wider than the indifferent. For although all indifferent states of affairs are neutral and all neutral states of affairs are the same in value, there will be some neutral states of affairs that are not indifferent: these will be the negations of good states of affairs and the negations of bad states of affairs. And so where Meinong would say that the negation of a good state of affairs is bad, or that the negation of a bad state of affairs is good, we say that it is neutral.

3.

This way of looking at intrinsic value throws light upon what G. E. Moore called "the principle of organic unities" – a principle he expressed by saying "the value of ... a whole bears no regular proportion to the sum of the values of its parts," and "the value of a whole must not be assumed to be the same as the sum of values of its parts." [11] The principle is readily illustrated in terms of what Meinong, in the *Ethische Bausteine* called "*Sekundärwerthaltungen*" – the ways in which one man may value the ways in which another man may share in value.[12]

We may suppose that joy is good and sorrow is bad. Consider now the joy or sorrow that one man may take in another man's joy or sorrow. There are these possibilities: *Mitfreude*, or joy in the other man's joy; *Mitleid*, or sorrow in the other man's sorrow; *Schadenfreude*, or joy in the other man's sorrow; and envy or *Neid* – sorrow in the other man's joy. Though Meinong is not clear about this point, let us take these terms intentionally so that the occurrence of these "*Sekundärwerthaltungen*" on the part of the one man will not imply the existence of the corresponding "*Primärwerthaltung*" on the part of the other. Saying that a man experiences *Mitfreude*, then, will be to say that he experiences joy in what he takes to be the fact that another man experiences joy, but it will not imply that anyone actually does experience joy. And analogously for the other terms: I may envy you for what I take to be your good fortune without it being the case that you actually have ever had good fortune.

Given an ethics such as that of Brentano, *Schadenfreude* and *Mitleid* provide clear cases of what Moore called organic unities: they are wholes having a value that is not proportional to the sum of the values of their parts.[13] We have assumed that joy is good and that sorrow is bad. Suppose, then, we go on to say that that joy which is joy in another man's sorrow is neutral and therefore neither good nor bad,

---

[11] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 27, 28.

[12] *Abhandlungen zur Werttheorie*, pp. 712-718.

[13] See the discussion of pleasure in the bad ("Lust am Schlechten") and displeasure in the bad ("Unlust am Schlechten") in Franz Brentano, *Ursprung sittlicher Erkenntnis*, Third Edition (Leipzig: Felix Meiner, 1934), pp. 84-86. This discussion may be found in the English edition, *The Origin of Our Knowledge of Right and Wrong* (London: Routledge & Kegan Paul, 1969), edited by Roderick M. Chisholm.

and that that sorrow which is sorrow in another man's sorrow is also neutral and therefore neither good nor bad. If, now, conforming to Moore's terminology, we call any state of affairs a "whole" and any state of affairs entailed by such a whole a "part" of that whole, we may say that *Schadenfreude* is a neutral whole that has a good part but no bad part. Taking pleasure in what one takes to be another man's sorrow has a good part – there being pleasure. It has no bad part – for, as we have seen, taking pleasure in what one takes to be another man's sorrow does not imply that there *is* such sorrow, and it does not imply the existence of any other evil. If the whole has a good part, and no bad parts and is nevertheless neutral, then surely it has a value that is not proportionate to the value of the sum of its parts. And analogously for *Mitleid.* If it has a bad part (there being sorrow), has no good parts, and is nevertheless neutral, then it, too, has a value that is not proportionate to the value of the sum of its parts.

Such "organic unities" are to be contrasted with what Moore called "mere sums" – wholes having a value that *is* proportionate to the sum of the values of their parts. An example would be that whole which is Jones taking pleasure in the being of stones and Smith taking displeasure in the being of stones. This whole has a good part and a bad part which balance each other off with the result that the whole is itself neutral. But the good part of that whole which is *Schadenfreude* is not balanced off by any bad part; yet the result is neutral. And the bad part of that whole which is *Mitleid* is not balanced off by any good part; yet the result is neutral. One might say that in the case of *Schadenfreude* the goodness of one of the parts is *defeated* by the larger whole and that in the case of *Mitleid* the badness of one of the parts is *defeated* by the larger whole.[14]

These concepts, I believe, are captured by the following definitions. We will say that an *Objektiv* or state of affairs p is *part* of an *Objektiv* or state of affairs q, provided only p entails q. Following Moore, we will say that a *mixed whole* is one having good parts and bad parts and an *unmixed whole* is one that is not mixed. (Given our assumptions, every whole, whether good or bad and whether mixed or unmixed, will have neutral parts.) Let us say that a state of affairs p *falls in value between* two states of affairs, q and r, provided that either (a) q is preferable to p and p is preferable to r or (b) r is preferable to p and p is preferable to q. We may now say that a state of affairs p is an *organic whole*, or an *organic unity*, provided only the following condition holds: either (a) p is a mixed whole and does not fall in value between any two of its parts or (b) p is unmixed and either has a good part that is better than p or has a bad part that is worse than p. Thus *Schadenfreude* is an organic unity, by this definition, since it is an unmixed whole (it has a good part but no bad part) and has a good part that is better than it; and *Mitleid* is an organic unity since it is an unmixed whole and has a bad part that is worse than it. On the other hand, that neutral state of affairs which is Jones taking pleasure in the being of

[14] For other examples and a more detailed discussion of defeat, see Roderick M. Chisholm, "The Defeat of Good and Evil," *Proceedings and Addresses of the American Philosophical Association*, Vol. XLII (1968-69), pp. 21-38.

stones and Smith taking displeasure in the being of stones is not an organic unity, for it is a mixed whole that falls in value between two of its parts. A state of affairs which is thus not an organic unity might be said to be a *mere sum*.

If a whole is an organic unity then it could be said to *defeat* – wholly or partially – the value of some of its parts. Thus *Schadenfreude* is a neutral whole which wholly defeats the goodness of one of its parts and *Mitleid* is a neutral whole which wholly defeats the badness of one of its parts. These concepts may be defined as follows.

The goodness of a state of affairs p is *wholly defeated* by a wider state of affairs q provided that: p is part of q, p is good, q is not good, and any part of q that is worse than q is part of p. And the badness of a state of affairs p is wholly defeated by a wider state of affairs q provided that: p is part of q, p is bad, q is not bad, and any part of q that is better than q is part of p.

Total defeat, thus defined, may be contrasted with partial defeat. The goodness of a state of affairs p may be said to be *partially defeated* by a wider state of affairs q provided that: p is part of q, p is good, p is better than q, and any part of q that is bad is part of p. And the badness of a state of affairs p may be said to be partially defeated by a wider state of affairs q provided that: p is part of q, p is bad, q is better than p, and any part of q that is good is part of p. (What would illustrate partial defeat? Consider a man's being justly *punished* for a wicked deed, according to the retributivist view: it entails that evil which is the performance of the misdeed; it contains a further evil which is the suffering and disgrace of the man who is punished; but this suffering and disgrace – according to retributivism – makes things better than they would otherwise have been. And so on this view the whole situation – the punishment of the man who performed the wicked deed – may be bad but it is better than that part of it which is just the performance of the wicked deed. Hence the badness of that performance is partially defeated; for the whole, though bad, is better than *it* is, and the whole, having no good parts at all, has no good parts that are not part of the wicked performance.[15])

Given these definitions and the principles set forth above, we may say that, if there are *no* organic unities, if every whole is a mere sum, then: (1) a whole is mixed if and only if it has two parts such that it falls in value between them; (2) if a whole is unmixed, then either it is at least as good as any of its good parts or it is at least as bad as any of its bad parts; and (3) no whole defeats – either wholly or partially – the goodness or badness of any of its parts.

[15] Thus Brentano held in an unpublished fragment that a bad state in combination with sorrow could be better than that same bad state in combination with pleasure, and cites vindictive punishment, repentance, and atonement as possible examples. The fragment, listed in Brentano's *Nachlass* as *Ethik 1*, is entitled "Vom Guten, das in der Zuordnung liegt."

4.

One further concept, suggested in Meinong's writings, may be of fundamental importance for the theory of value. The bearers of value, according to Meinong, are *Objektiva,* or *states of affairs.* Meinong also used the term "*Dignitiva*" for objects of feeling (and "*Desiderativa*" for objects of desire).[16] What distinguishes a *Dignitativum* from an *Objektivum*? I am not certain that I have understood Meinong, but it may be that he has the following concept in mind.

Let us say that an *Objektivum* or state of affairs p is a *V-state* provided only that p is a nonneutral state of affairs which is such that, for every q, if p entails q and q has the same value as p, then q entails p. A V-state is thus a minimal value state. Jones being happy is not a V-state, for it entails someone being happy which has the same value as it but does not entail it. It being the case both that Jones is happy and Jones' wife is happy is not a V-state, for it entails two people being happy which has the same value as it but which does not entail it. But presumably these are V-states: someone being happy, someone acting virtuously, someone being unhappy, it being the case the amount of unhappiness is greater than the amount of happiness.

It is possible that Meinong was thinking of V-states when he spoke of "*Dignitativa*". Perhaps the following will indicate why I believe the concept to be of fundamental importance for the theory of value.

We may say "the value of a state of affairs is a function of the value of the V-states it entails" and explicate this as follows: (1) Every nonneutral state of affairs p entails a set of V-states such that the conjunction of them all has the same value as p. (2) A state of affairs p is preferable to a state of affairs q, if and only if: either (a) the conjunction of all the V-states that p entails is preferable to the conjunction of all the V-states that q entails; or (b) p entails no V-states and the conjunction of all the V-states that q entails is bad; or (c) q entails no V-states and the conjunction of all the V-states that p entails is good. And (3) a state of affairs p has the same value as a state of affairs q, if and only if: either (a) the conjunction of all the V-states that p entails has the same value as the conjunction of all the V-states that q entails; (b) one entails no V-states and the conjunction of all the V-states that the other entails is neutral; or (c) neither entails any V-states.

If we apprehend the value of a V-state, we do so directly. One way of apprehending the value of a state of affairs which is not a V-state is this. We first determine what V-states it entails. If we find that it entails no V-states, we know that it is neutral. If we find that it does entail V-states, we then attempt to apprehend the value of that V-state which is the conjunction of all the V-states it entails. Is there a general formula by means of which the value of a disjunction, $p \vee q$, can be exhibited as a function of the value of its disjuncts? There is this: "Consider the worst in p and the worst in q, and take whichever one r is not worse than the other; consider the best in p and the best in q, and take whichever one s is not better than the other. Then: (1) $p \vee q$ has the same value as r&s; (2) if r is bad and s is neutral, then $p \vee q$ has

[16] See *Abhandlungen zur Werttheorie*, pp. 397, 401, 636, and Meinong's "Selbstdarstellung," p. 110.

the same value as r; and (3) if s is good and r is neutral, then p∨q has the same value as s." In applying this formula, we assume that every state of affairs implies some neutral state of affairs. Is there, analogously, a general formula by means of which the value of a conjunction can be exhibited as a function of the value of its conjuncts? It would seem not. For the conjunction may imply something about the distribution of value which is not implied by either conjunct separately; and the value of the conjunction may be a function of this distribution as well as of the amount of value in its conjuncts. On the whole conjunction may defeat the value of one of its conjuncts. Hence our only recourse is to determine what V-states the conjunction entails and then to try to apprehend directly the value of their conjunction.

There are many unanswered questions, then, in the theory of value and in the area of the logic of the emotions. I hope that those who work in the field will continue to follow in the footsteps of Brentano, Moore, and Meinong.[17]

[17] Prepared in part under NSF Grant GS-2953.

CHAPTER 14

N. M. LEMOS

THE BEARERS OF INTRINSIC VALUE

In this chapter, I consider what are the bearers of intrinsic value or what are the kinds of things that are intrinsically valuable. Are they abstract objects or concrete, particular things or both? If they are abstract objects, then are they properties, facts, or states of affairs? If they are particulars, are such things as persons, apples, and cars bearers of intrinsic value?

1. THE BEARERS OF VALUE: ABSTRACT OBJECTS

What are the bearers of intrinsic value? What are the kinds of things that have intrinsic value? Among the traditional candidates, we may distinguish between those that are abstract objects and those that are concrete, individual things such as persons, dogs, and cars. Let us begin by considering the former.

Concerning abstract objects, there are at least three main candidates: properties, states of affairs, and facts or states of affairs that obtain. Ordinary discourse sometimes suggests that properties are intrinsically good or bad. People sometimes say such things as "Pleasure and wisdom are intrinsically good" and "Pain is intrinsically bad." The view that some properties are intrinsically good has been defended by Panayot Butchvarov. Some philosophers, including Chisholm, have held that states of affairs are the bearers of value. Others have held that facts are bearers of value. This view has been defended by W. D. Ross, who writes, "what is good or bad is always something properly expressed by a that-clause, i.e. an objective, or as I shall prefer to call it, a fact."[1] In this section, I defend the view that facts or states of affairs that obtain have intrinsic value, whereas properties and states of affairs that do not obtain do not have intrinsic value.

In discussing the bearers of value, I make certain metaphysical assumptions about properties and states of affairs. I make further use of these concepts in Chapter 7 in connection with the "distinctiveness" of intrinsic value. Let us define "*x* is a

[1] W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930), p. 137.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 181-190.
© 2005 *Springer. Printed in the Netherlands.*

property" as "it is possible that there is something that exemplifies *x*."[2] According to this definition, we must say that there are no properties that cannot be exemplified. Thus, although there is a property of being round and a property of being square, there is no property of being a round square. I make the following metaphysical assumptions about properties. I assume that there are properties that exist and are exemplified, and that there are properties that exist but are not exemplified. For example, I assume that the property of being female exists and is exemplified but that the property of being a mermaid exists but is not exemplified. The view that there are properties that are not exemplified is sometimes taken to be a mark of *extreme realism*. I also assume that properties can be the objects of certain intentional attitudes, that they can be conceived, considered, and attributed. I assume that one can, for example, consider and attribute the property of being a mermaid, even if nothing has that property. For example, a drunken sailor with poor vision at dusk might attribute to a sea-weed-draped manatee the property of being a mermaid.

The following passage from Chisholm's *Person and Object* will help clarify the concept of a state of affairs that I employ:

> States of affairs, as they are being considered here, are in no way dependent for their being upon the being of concrete, individual things. Even if there were no concrete, individual things, there would be indefinitely many states of affairs. States of affairs, so conceived, resemble what have traditionally been called *propositions* in the following respect. Even though the author of *Waverly* was the author of *Marmion*, "the author of *Waverly* being knighted" expresses a different state of affairs than "the author of *Marmion* being knighted" (the former state of affairs but not the latter could obtain in worlds in which there is no *Marmion,* and the latter but not the former could obtain in worlds in which there is no *Waverly*).[3]

I make the following assumptions concerning states of affairs. First, I assume that there are states of affairs that exist but do not obtain, and that there are states of affairs that exist and do obtain. The state of affairs *everyone's being wise and virtuous* exists but does not obtain, and the state of affairs *someone's being happy* exists and obtains. Second, there are states of affairs that necessarily obtain and states of affairs that necessarily do not obtain. The latter may be said to be "impossible" states of affairs. An example of the latter is the state of affairs *there being something that is round and square.* (Of course, if what I've said concerning properties is correct, this state of affairs does not involve a property of being round and square. We may say instead that it involves the property of being round and the property of being square.) Third, I say that a fact is a state of affairs that obtains. Thus, we may say that someone's being happy is a fact, but everyone's being wise and virtuous is not a fact. Finally, I assume that states of affairs can be the objects of

---

[2] See Roderick Chisholm's definition of an "attribute" in *On Metaphysics* (Minneapolis: University of Minnesota Press, 1989), p. 100. Chisholm says, "*P* is an attribute = *Df. P* is possibly such that there is something that exemplifies it."

[3] Roderick Chisholm, *Person and Object* (LaSalle, Ill.: Open Court Publishing Co., 1976), p. 114.

intentional attitudes. They may be considered and contemplated, accepted or rejected, loved or hated, desired or preferred, and so on. The state of affairs *there being a snake in the grass* is such that one can consider it, accept it, and fear it.

Let us begin by considering the view that properties are bearers of value. In considering this claim, we may note that in ordinary discourse people say such things as "Pleasure is good for its own sake" and "Perfect justice is good in itself." But such talk is not a very good reason, if it is any reason at all, for accepting the view that properties as such are intrinsically good or bad. It may be that what people really mean is more accurately expressed by the more complicated locution "Someone's being pleased is good for its own sake" and "There being perfect justice is good in itself." Butchvarov's defense of the thesis that properties are bearers of value does not rest on dubious considerations about ordinary language. Instead he writes:

> We may say that a certain person's life is good, but we may also say that happiness as such, that is the property a life may have of being happy, is good. Now, I suggest, a person's life can be said to be good on the grounds that it is happy only if happiness itself can be said to be good, and in general a concrete entity can be said to be good only on the grounds that it has some other property or properties that themselves have the property of being good.[4]

We may surely agree with Butchvarov that (1) a life may have the property of being happy and (2) a person's life can be said to be good on the ground that it is happy. But does it follow from these assumptions that the property of being happy, as distinguished from the particular happy life, is itself good, that the property of being happy has the property of being good? I don't think so. In response, we should consider the following two points. First, we may recognize (a) that something can have the property of being colored and (b) that a thing can be said to be extended on the grounds that it is colored, but it does not follow from these assumptions that the property of being colored has the property of being extended. I assume that the property of being colored, unlike particular colored things, is not extended. In general, from the fact that X has G because X has F, it does not follow that the property of being F has the property of being G. Second, let us consider the parallel case of rightness. We may assume that right actions are right in virtue of their having certain properties. For the sake of argument, suppose that some actions are right because they have the property of maximizing intrinsic value. Must we say that an action is right only because it has some property that itself has the property of being right? I am not sure that this makes any sense, for right and wrong seem to be properties of actions and not of properties. If we say "Maximizing value is right," surely what this means is not that the property of maximizing value is right, but rather something to the effect that any action that has that property is right. If this is so, then Butchvarov's remarks provide no compelling reason for us to think of goodness as a property of properties.

---

[4] Panayot Butchvarov, *Skepticism in Ethics* (Indianapolis: Indiana University Press, 1989), p. 14.

I suggest that, strictly speaking, it is not pleasure or perfect justice considered as abstract properties that have intrinsic value. What has intrinsic value is surely more accurately thought to be the exemplification or possession of these properties, or, we may say, the *fact* that someone is pleased or that a community is perfectly just. A world containing nothing that exemplifies properties such as wisdom, beauty, or pleasure, or that has no concrete particulars at all, would be a world without anything intrinsically good, even if it contained the abstract, unexemplified properties of wisdom, beauty, or pleasure. But even if such properties are not themselves intrinsically good, we may say that properties such as pleasure, wisdom, and beauty, are "good-making" properties in the sense that the fact that something has them is intrinsically good.

Are facts bearers of value? I assume that the answer is "yes." If it is a fact that someone is suffering, then that fact is intrinsically bad; and if it is a fact that someone is happy, then that fact is intrinsically good. If facts are states of affairs that obtain and if facts are bearers of value, then there is an understandable temptation to say that some states of affairs are bearers of value. This is simply because facts are states of affairs that obtain. But for the sake of precision and clarity, it is useful to distinguish facts and states of affairs and to speak of facts as bearers of value.

What of those states of affairs that do not obtain, that are not facts; are they intrinsically good or bad? Do they have intrinsic value? There are considerations both for and against an affirmative answer, but I think the correct answer is "no." Let us consider the case for an affirmative answer. The weightiest reason for an affirmative answer arises from thinking of intrinsic value in terms of correct or fitting emotional attitudes. Suppose "*A* is intrinsically good" means "*A* is worthy of love" or "*A* is necessarily such that, for any *x*, the contemplation of just *A* by *x* requires that *x* love *A* as such." Whether a state of affairs is worthy of love or a pro-attitude does not depend on whether it obtains or is a fact. We can say, correctly I think, that someone's being perfectly honest is worthy of love even if no one is perfectly honest. Similarly, we may say that it is fitting to prefer everyone's being very happy to everyone's being slightly happy even if no one is happy. Therefore, if *A* is intrinsically good just in case *A* is worthy of love, then *A* can be intrinsically good even if it is not a fact.

Against this view, however, is the main consideration raised against the view that properties are bearers of value, namely, that what seems to be intrinsically good is the actual exemplification of certain good-making properties such as wisdom or pleasure. What seems to be intrinsically valuable is that something *has* these properties, that a certain state of affairs involving these properties obtains. Again, I would urge that a world in which nothing exemplified such good-making properties as beauty, wisdom, or pleasure would be a world that contained nothing that is intrinsically good, even if it contained the nonobtaining state of affairs *someone's being wise and happy*. That intrinsic value requires the exemplification of certain sorts of properties is strongly suggested in the following passage from Ross:

> However much one were convinced that conscientiousness, for example, is good,... no one would say "*A*'s conscientiousness is good" if he were

> convinced that *A* is not in fact conscientious.… We might say "perfectly conscientious action is good" even if (as Kant suggests) we are not convinced that there ever has been such an action. But this is only a shorthand way of saying that without being sure that such an action has ever existed, we can be sure that if any existed it would be good. Hypothetical goodness presupposes hypothetical existence just as actual goodness presupposes actual existence.[5]

If it is true that states of affairs that do not obtain are not intrinsically good or bad, then we might say, speaking very loosely, of the states of affairs *everyone's being wise and happy* and *everyone's being perfectly just* not that these states of affairs *are* good, but that they *would be* good if they were to obtain. But even this is not quite right, for it would not be, strictly speaking, the state of affairs *everyone's being wise and happy* that would be intrinsically good, but the *fact* that everyone was so. Perhaps we could say that it would be good if the state of affairs were to obtain. Furthermore, if some nonobtaining states of affairs are worthy of love, then we should say that being intrinsically good implies being worthy of love, but being worthy of love does not imply being intrinsically good.

There is an objection to the view that facts are the bearers of value that is worth considering. As we have seen in the last chapter [* pp. 23-24 of this volume], both Chisholm and Moore accept the thesis of universality. We may take this thesis to imply that if *A* is intrinsically good, then it is not possible for *A* to exist and not be intrinsically good. Now consider the following argument. (1) If *p* is a contingent state of affairs, then *p* is a fact in some possible worlds but not in others. (2) Suppose the state of affairs *p* is intrinsically good in those worlds in which it obtains. (3) There are possible worlds in which the state of affairs *p* exists but does not obtain. (4) In those worlds *p* is not intrinsically good. (5) Therefore, *p* is intrinsically good in some worlds, but not in others in which it exists and does not obtain. The conclusion of this argument conflicts with the thesis of universality, since it implies that the intrinsic value of a thing is not essential to it. Should we abandon the thesis of universality on the basis of this argument? I do not think so. The argument assumes in the second premise that states of affairs rather than facts are the bearers of value. I have argued that it is the *fact* that *p* that is intrinsically good. If we are careful to distinguish between facts and states affairs, we may say that although the state of affairs that *p* can exist in possible worlds where it does not obtain, the fact that *p* does not exist in those worlds. There is no fact that *p* in those worlds where *p* does not obtain. Thus, in a world where *p* does not obtain, we cannot say that the fact that *p* exists in that world and is not intrinsically good. If we are careful to distinguish between states of affairs and facts, then we can say: if the fact that *p* is intrinsically good, then it is not possible for the fact that *p* to exist and not be intrinsically good. In other words, we can still hold that it is impossible for the fact that *p* to exist without being intrinsically good.

If the preceding remarks are correct, then, strictly speaking, we should not say that those states of affairs that do not obtain, that are not facts, *are* intrinsically good

---

[5] Ross, *The Right and the Good*, pp. 96-97.

or bad. We may say of certain states of affairs that do not obtain that they are such that it would be good for them to obtain or bad for them to obtain. Similarly, we should not say that one state of affairs is better than another unless they both obtain. Admittedly, in ordinary discourse, we often say with respect to states of affairs that do not obtain, and in some cases where both cannot obtain, that one is better than the other. Ordinarily we might say, for example, that Socrates's having exactly ten units of pleasure is better than Socrates's having exactly five units of pleasure. But if neither state of affairs obtains, then, strictly speaking, we cannot say that one *is* better than the other. Could we say that Socrates's having exactly ten units of pleasure and Socrates's having exactly five units of pleasure are such that if they were to obtain, the former would be better than the latter? This is problematic, since these states of affairs cannot both obtain at the same time. Instead let us say, somewhat more awkwardly, that there is some state of affairs, *q*, such that if *q* and Socrates's having exactly ten units of pleasure were to obtain, then Socrates's having exactly ten units of pleasure would be better than *q*, and if *q* and Socrates's having exactly five units of pleasure were to obtain, then *q* would be the same in value or better than Socrates's having exactly five units of pleasure.

2. THE BEARERS OF VALUE: CONCRETE PARTICULARS

At this point, I wish to turn from consideration of abstract objects such as properties, facts, and states of affairs and discuss whether certain concrete, particular things are also bearers of value. The concrete particulars I wish to consider are individual things such as human beings, dogs, apples, and cars. These things are not abstract objects; they are not properties, facts, or states of affairs. I argue that such concrete particulars are not bearers of intrinsic value.

We often judge that such things as apples, people, and dogs are good or bad. We judge that a particular apple or car is good or bad. But clearly, some judgments of value are not judgments of intrinsic value, and it hardly follows from the fact that a person or an apple is good in *some* sense that it is *intrinsically* good. Such things as apples, cars, and human beings might have instrumental value, but if they are good in this sense, it does not follow that they have intrinsic value. Particular apples, cars, and people may also be good in the sense of being *good members of their kind.* But the fact that something is a good member of its kind does not imply that it has intrinsic value. If we agree that a particular knife can be a good knife or that a particular toothpick can be a good toothpick, without either one having intrinsic value, then clearly, being a good member of a kind does not imply being intrinsically valuable.

I must note that I shall not consider here whether individual actions, events, or lives are intrinsically good. My reason for not considering these latter sorts of things is that there is some question of whether they are concrete particulars or whether they are more properly construed as abstract objects, as either species of properties, states of affairs, or facts, or akin to such things. We might, for example, think of an event or an action that occurs as a state of affairs that obtains. Thus, the occurrence

of an event such as something's falling, and the occurrence of an action such as someone's walking, might both be understood in terms of the obtaining of certain states of affairs. Trying to decide whether events are abstract objects or concrete particulars would take us too far afield. Instead, let me say simply that I have no doubt that some actions, events, and lives are intrinsically good, and if it is reasonable for us to believe that such things are concrete, particular things, then it is reasonable for us to believe that some concrete particulars are intrinsically good and that others are intrinsically bad.

Let us consider, then, whether such things as human beings, apples, dogs, and cars can be bearers of intrinsic value. It is not at all clear that such concrete, individual things are bearers of intrinsic value. In considering whether properties and states of affairs are bearers of value, it was urged that it is not they that have intrinsic value, but rather the exemplification or possession of certain properties or the obtaining of certain states of affairs. If this is right, then it seems at least prima facie plausible that it is not a dog, an apple, or a person as such that has intrinsic value but rather such things as a dog's being pleased, a person's being wise, or even, perhaps, the fact that there are persons and dogs.

Among those concrete individuals that may seem most plausibly to be bearers of value are persons and human beings. If persons and human beings do not have intrinsic value, then it is doubtful that such value is had by any concrete particulars. Now as with cars and apples, the mere fact that persons and human beings are good in some sense does not imply that they are intrinsically good. The fact that John is a good painter or a good thief does not imply that John is intrinsically good any more than the fact that *X* is a good pencil implies that *X* is intrinsically good. Even the fact that John is morally good does not imply that *he* is intrinsically good. We may note that even if the fact that John is morally good is itself intrinsically good, it does not follow that *John* is intrinsically good. In general, if someone's having property *F* is intrinsically good, it does not follow that either (1) the person who has *F* or (2) the property of *F* itself is intrinsically good. On the other hand, it does seem reasonable to hold that if some concrete, individual, *A*, *were* intrinsically good, then the fact that *A* exists *would be* intrinsically good.

The existence of persons does seem to be necessary for the existence of some important intrinsic goods. If, for example, the properties of being wise and having morally good emotions can only be had by persons, and if the having of these properties is intrinsically good, then the existence of persons is necessary for their existence. But the fact that persons are necessary for the existence of such goods does not imply that persons are themselves intrinsically good. The existence of such goods depends just as much on the existence of the properties of being wise and having morally good emotions. Since we have already seen that properties are not bearers of value, the fact that persons are necessary for the existence of such goods does not imply that they are bearers of intrinsic value.

I am skeptical that persons are bearers of intrinsic value. Imagine a world in which nothing is beautiful or wise or morally good, a world that contains no pleasure or pain, but contains persons in a deep, dreamless sleep, their bodies tended

by machines, a world in which the only living things are these sleepers. Now, such a world might perhaps be thought to have a greater potential for the exemplification of intrinsic value than an otherwise similar world that lacked any persons at all. One might make such a claim thinking that there was a greater potential in the former for the realization of such things as moral goodness, wisdom, and pleasure. (Of course, the potential for moral wickedness, stupidity, and pain might be just as great.) But even if that were true, such a potential would be irrelevant to the assessment of the world's intrinsic value, since the fact that one state of affairs or fact is more likely to lead to the exemplification of intrinsic value than another implies nothing about the intrinsic value of either state of affairs or fact. When we consider the question of intrinsic value, such a world seems to be without any intrinsic value; it does not seem to be a world that anyone should desire or favor in and for itself. The mere fact that there are persons in that world contributes no intrinsic goodness to it. And the mere fact that there are persons does not seem to be desirable in and for itself or intrinsically good. But if the fact that there are persons is not intrinsically good, then it seems reasonable to believe that persons are not intrinsically good, since if individual persons had intrinsic value, the fact that they exist would be intrinsically good.

On reflection, I suggest that what we find intrinsically good are not concrete, individual persons or the mere fact that they exist, but such things as their being wise or happy or morally good, states of affairs that do not obtain in our imagined world. If persons and human beings as such are not intrinsically good and if they are the most plausible candidates among concrete individuals for bearers of value, then we have reason to doubt that any concrete, individual things are bearers of value.

Finally, let me advance one last set of considerations against concrete, individual things as bearers of value. In asking whether such things are bearers of intrinsic value, we should consider what is involved in the attitudes of love, hate, and preference and the related attitudes of desiring and wanting. There is reason to believe that the objects of such attitudes are not simply concrete, individual things, such as a dog, an apple, or a person, but more complex sorts of things, such as states of affairs, facts, and perhaps events. In ordinary contexts, when someone says, "I want a yacht," what he wants, the object of his want, is not simply a yacht, but more precisely his owning or having a yacht. Similarly, if someone says, pointing to a piece of apple pie, "I want that," typically what he wants is to eat it or taste it. In this case, the object of his want or desire, what he wants or desires, is not merely the piece of pie. When people say such things as "I desire *X*" or "I want *X*," where *X* is some concrete, individual thing, we must often gather from the context the object of their want or desire. If Mary says, "I want a little lamb," she may mean that she wants to own a lamb, or to play with a lamb, or to have a little more lamb on her plate. We can usually tell in a given context what is wanted or desired. But what I chiefly wish to emphasize is that the object of the want or desire always seems to be

an object of a more complex and different sort than the mere concrete, individual thing.[6]

This feature of our wants and desires appears to belong to our preferences as well. If someone says, "I prefer *X* to *Y*," where *X* and *Y* are concrete, individual things, what is more accurately said to be the object of his preference is not merely *X* and *Y*, but some other, more complex sort of thing. Thus, if someone says he prefers a piece of chocolate cake to a piece of apple pie, the object of his preference might be more precisely said to be his eating or tasting chocolate cake to his eating or tasting apple pie. Similarly, if someone says he prefers Mary to Jill, his attitude might be more accurately said to be his preferring to be in Mary's company to Jill's or to have Mary as a boss to having Jill as a boss or to read Mary's books to reading Jill's.

The view that the objects of desires, wants, and preferences are not merely concrete individuals may be further supported by reflecting on the peculiarity of supposing that someone should want or desire a concrete individual without wanting or desiring some more complex object that somehow involved that thing. Think how strange it would be were someone to insist that he wants a yacht while denying that he wanted to own a yacht or that he wanted a yacht to sail, steer, see, touch, care for, admire, run aground, and so on. It does not seem possible that anyone can merely or simply want or desire a concrete, individual thing.

These remarks about the objects of emotional attitudes such as desiring and preferring are relevant to the question of whether concrete, individual things are bearers of intrinsic value. For if what we have said about the objects of preference is correct, if they are not mere concrete individual things but more complex objects, then the sorts of things that it can be fitting or appropriate to prefer must be objects of a complex kind, different from such things as a dog or a human being. Consequently, if we assume that whatever is intrinsically good is preferable to whatever is intrinsically bad, then both what is intrinsically good and what is intrinsically bad must be possible objects of preference. But if what is a possible object of preference is not merely a concrete individual such as a dog or a human being, then such things are not, strictly speaking, intrinsically good or intrinsically bad; they are not bearers of intrinsic value. If we assume, therefore, that whatever is intrinsically good is preferable to whatever is intrinsically bad, then the sorts of things that are intrinsically good or bad must be complex objects like states of affairs or facts.

The preceding remarks offer some support for the view that concrete, individual things are not bearers of intrinsic value. Some philosophers have argued that what has been said about the objects of desiring, wanting, and preferring also pertains to the objects of love and hate. According to this view, the objects of love and hate are not mere concrete, individual things but more complex objects. There are several forms that this view might take. For example, one might hold that loving and hating

---

[6] Ibid., p. 96: "But obviously it is only a rough and ready description of my desire to say that I desire a primrose. What I desire is to be seeing it or smelling it or possessing it."

concrete, individual things can be reduced to the phenomena of preference. In its simplest form, this view would claim that loving *X*, where *X* is a concrete, individual thing, is a matter of preferring the state of affairs *X's existing* to any state of affairs to which one is indifferent. (One is indifferent to a state of affairs when one does not prefer it to its negation and vice versa.) Thus, to say that I love my old rocking chair is to say that I prefer my old rocking chair's existing to any state of affairs to which I am indifferent. If the phenomena of love and hate can be reduced to that of preference, then the objects of love and hate, like that of preference, would seem to be complex objects like states of affairs and facts.

An alternative view might claim that the objects of love and hate are not mere concrete individuals, without reducing love or hate to preference. To love someone, according to this view, is typically, though not always, a matter of loving such things as his being or existing, his being healthy, his being happy, to see him and to spend time in his company. It is to love a constellation, and possibly a shifting variety, of complex objects. This view about the objects of love is expressed in the following passage by Everett Hall:

> "John loves Mary" would then need to be interpreted as elliptical. What John loves is Mary's having a pair of blue eyes, a dimple in her right cheek, a way of turning suddenly pensive in the midst of a playful mood, a ..., a ..., etc. No one of the characterizations is sufficient (unless John is a very, very simple boy); it is only the ensemble that is the object of his passion. It would follow that when John tells Mary, "I love you", he is saying something very complex, indeed... I am suggesting...that a complete description of an emotion must, besides the emotive verb, contain a secondary one subordinate to it, either in an oblique participial phrase or in a substantive clause in the accusative.[7]

According to this view, the object of love is not merely the concrete, individual thing, nor is it a mere bundle of properties or characteristics. John's loving Mary is not merely a matter of his loving the property of having blue eyes, a dimple, and so on. The object of John's love is also not merely some existential generalization such as someone's having blue eyes or someone's existing. The object of John's love is not simply Mary but some more complex objects that, to speak very loosely and roughly, involve Mary.

I am not entirely sure that either of these views about the objects of love is correct. Yet if either is correct, we have one more reason to think that the bearers of intrinsic value are not concrete, individual things. For if *p* is intrinsically good implies that *p* is worthy of love, then *p* must be the sort of thing that can be loved, *p* must be a possible object of love. But if the preceding remarks are correct, and the objects of love and hate are not concrete, individual things, then such things are not bearers of value. If mere concrete, individual things are not the objects of love or hate or preference, then such things are not bearers of intrinsic value.

---

[7] Everett Hall, *Our Knowledge of Fact and Value* (Chapel Hill: University of North Carolina Press, 1961), p. 147.

CHAPTER 15

M. J. ZIMMERMAN

INTRINSIC VALUE AND INDIVIDUAL WORTH

The headline in this morning's *Daily News* proclaims, in block letters 5 cm tall:

DIANA'S DRESS CAUSES BIG SENSATION!

We all understand what it means, but should any of us take it literally?

What the headline means is that there was something about Diana's dress that caused a big sensation. The headline is intended to entice us to read further and discover what this something was. (Was the dress especially lavish? Was it shockingly revealing? Was its design outrageous? What exactly was it that caused such a stir?) Once we have learned these details, we will have a better understanding of what took place.

Suppose the dress was especially lavish, carrying a price tag of £30,000. It is this that caused the sensation. Once we have discovered this, what will our attitude be toward the claim made in the headline? Will we accept it as literally true? Will we, that is, still want to say that *the dress* caused the sensation, once we have learned that *the dress's being lavish* caused the sensation? Perhaps, but, if so, we surely wouldn't want to say that these causes are on a metaphysical par; for that would put us at risk of having to say that the sensation was causally overdetermined, which (we may assume) it was not. If we are not to be eliminativists of a certain sort and deny that the dress was, literally, a cause of the sensation, we must at least be reductionists and say that its being such a cause was nothing above and beyond some state of the dress being a cause of the sensation. Object-causation, if there is such a phenomenon at all, is metaphysically parasitic on state-causation, and so talk of the former is reducible to talk of the latter.[1]

The next morning's headline declares:

[1] Sometimes it might be better to say that events, rather than states, are causes. (Perhaps Diana's dress ripped as she emerged from her limousine, and this is what caused the sensation.) For the sake of simplicity, however, I will continue to talk in terms of states only. Note that I do not wish to deny that there might be an exception to the claim that talk of object-causation is reducible to talk of state-causation; perhaps agents are sometimes causes in such a way that their causation is not to be explicated wholly in terms of state-causation. My claim is only that the reduction is correct in the case of inanimate objects such as Diana's dress.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 191-205.
© 2005 *Springer. Printed in the Netherlands.*

DIANA'S DRESS OF GREAT VALUE!

Is this something we should take literally?

That depends on the sort of value at issue. If it is economic value, then it seems quite natural to take it literally (and also to accept it as true; £30,000 is a lot of money for a dress). But suppose that this headline appears, not in the *Daily News*, but in the *Axiological Gazette*, and that the value at issue is intrinsic value. Then, I believe, either we should not take the headline literally, or we should understand the value at issue to be metaphysically parasitic on the intrinsic value of states.

1.

To defend this position, I must say what I mean by "intrinsic value."

Often, when introducing students to the notion of intrinsic value, I set it in the context of hedonism. I don't try to define the notion, merely to illustrate it.

"Is charity a good thing?" I ask.

"Yes," comes the reply.

"Why?"

"Because it provides people with food, clothing, and shelter."

"What's good about that?"

"It satisfies their needs."

"What's good about that?"

"It gives them pleasure."

"What's good about that?"

At this point, some of the students try to think of something else that might in turn explain the value of pleasure, but I tell them that it is here that the hedonist puts a stop to this line of inquiry by saying simply, "It just is." And then I elaborate by telling them,

"What the hedonist is saying is that charity is (usually) good because of what it produces, which is, ultimately, pleasure, whereas pleasure is good, not because of what it produces, but rather because of what it is. That's the difference between intrinsic value and instrumental value."

I then provide a second, briefer illustration.

"Is hitting someone on the head with a hammer bad?"

"Yes."

"Why?"

"Because it causes pain."

"What's bad about that?" I ask rhetorically, and continue, "This is where the hedonist would stop. He'd say that pain just is bad. It's bad because of what it is, not because of what it produces."

Of course, this is all very rough, but the students seem to get the idea. Many philosophers, however, would accuse me of misleading my students. They'd claim that to say that pleasure is good "because of what it is" and that pain is bad "because of what it is" is to say nothing helpful, since it provides no insight into the type of

value at issue. They'd also say that I am confusing two types of value that Christine Korsgaard has famously argued must be kept distinct: first, there is intrinsic value, which is the value that something has "in itself" and which is to be contrasted with extrinsic value; then there is final value, which is the value that something has "for its own sake" and which is to be contrasted with instrumental value.[2] The charge is that, in contrasting intrinsic value with instrumental value, I am confusing intrinsic value with final value.

I shall try to explain why both objections are mistaken, since states, and only states, are the bearers of intrinsic value.

2.

Let me begin by conceding that the crucial value at issue is that which Korsgaard calls final value. It is the value that something has for its own sake that is the ground of all attributions of value (or, at least, all attributions of the sort of value I'm concerned with here[3]). Consider the claim that pleasure is good for its own sake. I take this to mean that every state of pleasure, every state consisting of someone's being pleased, is good simply in virtue of being such a state. There is no helpful explanation why the state is good; it just is good "as such," that is, good in virtue of its own nature. But though unhelpful, in that this account of the goodness of pleasure does not cite something else in terms of which the goodness of pleasure may be understood, it does, contrary to the first objection, provide some insight into the nature of final value. Such value is *nonderivative*; it is the ground or source of nonfinal values (such as those of charity and of hitting someone on the head with a hammer), values that may thus be declared derivative.[4] All explanation must come to an end somewhere; the explanation of values stops with the citing of final values.[5]

---

[2] Korsgaard (1983), pp. 169-70; (1996), p. 250 [* p. 78 of this volume].

[3] The sort of value at issue is ethical. Just how and why this is so is a topic that I cannot tackle here, however.

[4] Korsgaard uses the term "source" differently and, to my mind, rather oddly. In her (1983), p. 170 and (1996), p. 250 [* p. 78 of this volume], she says that something that has extrinsic value "gets [its value] from some other source," which suggests that something that has intrinsic value is the source of its own value. She thus thinks of "source" in terms of the intrinsic-extrinsic distinction rather than the final-nonfinal distinction.

[5] There is a complication. A distinction can be drawn between "basic" and "nonbasic" value. Suppose that Peter's being pleased is good for its own sake, and that Paul's being in pain is bad for its own sake. Presumably the compound situation of Peter's being pleased and Paul's being in pain may be assigned some value in light of the final values assigned to its components. What *kind* of value is it that this compound situation has? It seems natural to say that this, too, is final value. But it also seems to be derivative, in some way, from the values of its components. If so, then not all final value is nonderivative; only basic final value is.

It may be the idea that all final value is nonderivative that leads Moore to say that what is good for its own sake (or "ultimately good," as he sometimes puts it) contains no part that is not intrinsically good, whereas what is intrinsically good may contain such a part. (See Moore (1912), p. 31.) But this still doesn't quite capture the idea of something being nonderivatively good, since of course a compound may be comprised only of good components and thus apparently qualify as being good for its own sake, in Moore's sense.

Consider, now, the claims about final value that are sometimes made when it is individual objects, rather than states, that are said to have such value. Korsgaard, for instance, suggests that mink coats, handsome china, and gorgeously enameled frying pans might all be good for their own sakes.[6] Shelly Kagan claims that the pen used by Abraham Lincoln to sign the Emancipation Proclamation is something that might well be good for its own sake.[7] And Wlodek Rabinowicz and Toni Rønnow-Rasmussen are prepared (for the sake of illustration) to declare Princess Diana's dress good for its own sake.[8] If we were to ask these authors why the objects in question are good, we might expect, given what was said in the last paragraph, that they would simply answer, "They just are good 'as such.' They're good in virtue of their natures." But this is *not* what they say. Korsgaard attributes the value of the objects she mentions to their "instrumentality," that is, to their helpfulness in allowing us to accomplish certain tasks. Kagan attributes the value of Lincoln's pen to the unique historical role it played. Rabinowicz and Rønnow-Rasmussen attribute the value of Diana's dress to the fact that it belonged to Diana.

Notice two things about these claims made by Korsgaard *et al.* First, it is precisely because of such examples that they insist on the distinction between intrinsic value and final value; for the final values they claim on behalf of the objects they cite supervene on *extrinsic*, relational features of the objects.[9] And this brings me straight to the second point, which is simply that the sort of claim made by Korsgaard *et al.* is very different from the sort of claim made earlier about pleasure. In the case of pleasure, there was no helpful explanation why pleasure is good, and this fact fits well with the observation that the value of pleasure is nonderivative. But if we explain the values of the coat, the china, the pan, the pen, and the dress by appealing to certain particular relational properties of these objects rather than to their own natures, this seems to me a strong indication that these values are *derivative* and that we must press our inquiry further in order to reach the nonderivative values that are their sources.

A strong indication, I say, but naturally this can be contested. My contention is that a helpful explanation as to why something is good is available if and only if the goodness is derivative. I have also suggested that, where no helpful explanation as to why something is good is available, this is because the thing in question just is good "as such," that is, good in virtue of its own nature. Given that something's nature is intrinsic to it, my contention and suggestion jointly imply that *nonderivative value is intrinsic to its bearer*. Each of the assumptions on which this conclusion is based could of course be challenged, but further support for my position is provided by the observation that, given the particular relational properties cited by Korsgaard *et al.*,

---

[6] Korsgaard (1983), p. 185; (1996), p. 264 [* p. 89 of this volume]. She fails to distinguish adequately between something having value in the sense that it is valued and something having value in the sense that it is valuable. It is the latter that is at issue in this paper.
[7] Kagan (1998), pp. 285-6 [* p. 104 of this volume].
[8] Rabinowicz and Rønnow-Rasmussen (1999), p. 41 [* p. 121 of this volume].
[9] Kagan advocates continuing to use the term "intrinsic value" to refer to final value, but he insists on the distinction nonetheless.

we don't have to look very far at all in order to find the sources of the values they claim on behalf of the individual objects in question. It is the coat's (the china's, the pan's) being helpful that Korsgaard should identify as having final value; it is the pen's playing the historical role (or, at least, the sort of historical role) that it played that Kagan should say is of final value; it is the dress's belonging to Diana (or, at least, having some such connection to someone of that sort) that Rabinowicz and Rønnow-Rasmussen should claim to be of final value. The attribution of value to the objects is thus eliminable in terms of, or at least reducible to, an attribution of final value to some states of the objects. Just as in the case of causation, so too in the case of value: once the move from objects to states of objects is made, we get a fuller account and a better understanding of what is at issue.

3.

The claims that object-causation is parasitic on state-causation and that object-value is parasitic on state-value are, as I have already acknowledged, contestable. Nevertheless, I submit that the thesis is, in each case, an attractive and plausible one. But it is also a controversial one. I won't have much more to say regarding the thesis about causation, since that is not the concern of this paper. Rather, let me now address some objections to the thesis about value.

First objection

If something is valuable, this is because of some other property that it has. Given this, the principle upon which I am relying is unacceptable, for it leads to an infinite regress. The principle is this: if x is valuable because it has some property P, then x's having P is valuable. Now, if x's having P is valuable, this will be because of some further property Q that *it* has. The principle then implies that x's-having-P's having Q is valuable, which will in turn imply, for some property R, that x's-having-P's-having-Q's having R is valuable, and so on.

Response

I am not relying on the principle cited. This principle is in fact ambiguous, since "because" admits of more than one interpretation. Suppose that "because" expresses supervenience. Then I grant that, if something is valuable, this is because of some other property that it has. But I deny that its having this other property must be something that is itself valuable. Indeed, when it comes to final value, I think it is true that, if x (which, in my view, must be a state) is valuable because it has P, then x's having P is itself *not* valuable.

Suppose that I agree that Diana's dress is of great value because it belonged to Diana. I am *not* thereby saying that it has a certain final value that supervenes on the property of having belonged to Diana. That is what my *opponents* are saying. I am saying that it has a certain *non*final value that is related in some *other* way to this

property. (This nonfinal value will, like all value, supervene on some property, but it won't be the property of having belonged to Diana.[10] More on this in just a moment.) As I am using it, "because" in this case expresses not supervenience but derivation; and of course, when value is derivative in the present sense, it is derivative from something else that has value. My claim is that Diana's dress derives its value, not from the property of having belonged to Diana,[11] but from the state that consists in its having this property, and it is this state that has final value. (Actually, this is too simple. If there is final value to be found in this case at all, I'm sure it won't be, at bottom, in Diana's dress having belonged to Diana but rather, as indicated earlier, in something more general, such as Diana's dress having a certain sort of connection to someone of a certain sort.) The nonfinal value that Diana's dress has will thus supervene, not on the property of having belonged to Diana, but rather on the property of being a constituent of the state in question.

Second objection

If there is derivation of value, it will go in just the opposite direction from that which I have indicated. Suppose that it is a good thing that Diana's dress exists. This will be precisely because Diana's dress is valuable. To embrace reduction from the value of the object to the value of the state is "to put the cart before the horse… [I]t is the state that derives its value from the object it involves and not the other way round."[12]

Response

The state identified in this objection is not the state identified by me. Whether we call the "thing" in question, namely, that Diana's dress exists, a state or something else (a fact, for example), and whether we say that it is the sort of thing that can be good, it is not the sort of thing that I have called a state and which I have claimed to be the bearer of final value. The state that consists in Diana's dress having belonged to Diana is quite distinct from the fact that Diana's dress exists. Even if the latter has only derivative value, it doesn't follow that the former does. Indeed, insofar as it is plausible to say that the fact that Diana's dress exists has a value that it derives from

---

[10] One might doubt whether what I have called nonfinal value is really a type of value at all. Why say that something has a nonfinal value simply in virtue of being related to something else that has final value? (See the opening paragraph to Rønnow-Rasmussen (2002).) It's a good question, having to do with the distinction between eliminativism and mere reductionism. If one is to embrace the latter, one must say something about the *kind* of relation in question, since certainly not everything that is related in some way to something that has final value can be said therefore to have a value of its own; for then *everything* would be valuable. I won't pursue the matter here, however.

[11] In my view, properties, understood as Platonic entities capable of instantiation, are not the sort of thing that can have final value, being on the wrong side of the abstract-concrete divide.

[12] Rabinowicz and Rønnow-Rasmussen (1999), p. 43 [* p. 122 of this volume].

Diana's dress, it seems just as plausible to say that it derives this value from the dress's having belonged to Diana.

Third objection

To say that something is valuable is to say that it is the fitting object of a pro-attitude. Pro-attitudes vary with respect to the sorts of things to which they are directed. Certainly, states (and entities of that ilk, such as facts, propositions, and so on) can be the appropriate objects of pro-attitudes (such as preference and desire), but so can individual objects be the appropriate objects of pro-attitudes (such as love, admiration, and respect). Insofar as this is the case, it is individual objects that are the bearers of value.[13]

Response

I accept the general thesis that to be valuable is to be the fitting object of a pro-attitude.[14] (This can of course be challenged, but I think it contains an important insight into the nature of value.) And I also accept that individual objects can be the things toward which certain pro-attitudes are directed. Nonetheless I deny that individual objects can be the bearers of final value.

Notice that in many cases the attitudes that are directed toward individual objects may also be directed toward other things as well. I may love, admire, and respect someone, but I may also love, admire, and respect what he does. The fact that I have these pro-attitudes toward him doesn't itself show that he has final value, for these attitudes may derive from the attitudes I have toward what he does. I may admire what John does for its own sake (for example, I may admire his display of courage), and I may thus admire him for what he does (I may admire him for his display of courage), but this doesn't show that I admire John for his own sake.

Love sometimes differs in this respect from admiration, though. If you ask me why I love Kath, I may reply that it's because of her quick wit and curvaceous figure. Am I thereby indicating that I love her having this wit and figure? Yes. Does this mean that the love I have for her is merely derivative? No. The *sort* of love I have for her is different from the sort of love I have for her wit and figure. The love I have for her is romantic; the love I have for her wit and figure is not. While I nonromantically love her having the wit and figure that she has for its own sake and thereby have a derivative nonromantic love for her, this leaves untouched the fact that I do indeed romantically love her for her own sake.

Still, this is not enough to establish that Kath has final value. What is the link between her wit and figure, on the one hand, and my romantic love for her, on the

[13] Cf. Rabinowicz and Rønnow-Rasmussen (1999), pp. 46-7 [* p. 125 of this volume], and (2003) [* pp. 213-26 of this volume].

[14] Proponents of this thesis include, among many others: Brentano (1969), p. 18; Broad (1930), p. 283; Ross (1939), pp. 275-6; Ewing (1948), p. 152; Lemos (1994), pp. 12 and 15 [* pp. 20 ff. of this volume].

other? Perhaps it is merely causal. If so, this does nothing to show that my (romantic) love for her (for her own sake) is fitting or appropriate, and so we cannot conclude that she has final value.

But isn't my love for Kath, in light of her wit and figure, perfectly appropriate? Certainly, if all this means is that her having this wit and figure gives me *no* reason *not* to love her. But I take it that that is not all that's intended by the thesis at issue. Rather, the claim is that to say that something is valuable (for its own sake) is to say that one has reason *to* adopt a certain pro-attitude toward it (for its own sake, given that one is contemplating it in the first place). Thus, in the present context, the claim is that Kath's having the wit and figure that she has gives me reason to love her for her own sake. I think this is false.

Suppose that Kath has an identical twin, Kay, identical not only in terms of appearance but also in terms of personality, and so on. If Kath's wit and figure give me reason to love her for her own sake, then Kay's wit and figure give me equally good reason to love *her* for her own sake. But this is a disturbing thought.[15] It's disturbing because it seems incompatible with the attachment that I have to Kath in particular.[16] The fact is that neither Kath, nor Kay, nor I believe that reason somehow requires me to love Kay as I do Kath. Romantic love is simply not subject to duplication in this way. In this sense, it is not a rational attitude. For confirmation of this claim, compare romantic love with the sort of love that a stamp-collector may have for a certain stamp.[17] Suppose that it is the stamp's rarity (and nothing else – not its color, or shape, or whatever) that moves him. If his love for the stamp is to be rational, then surely he must love (or be prepared to love) any equally rare stamp (indeed, any equally rare object) just as much; otherwise, I cannot see how the emotion in question can be declared rational. This is important because it shows, I think, that it is *not* the stamp that is the immediate object of the collector's love. Nor, I should add, is it the stamp's *existence* that he immediately loves. Rather, it is the stamp's *being rare* that he immediately loves. It is its rarity that he immediately values, and that's why he values it in turn.

It might seem that, in saying this, I have overlooked the phenomenon of rational satiation. That I rationally want a piece of cheesecake doesn't imply that I will also want a second, qualitatively identical piece of cheesecake. Similarly, that I rationally love Kath doesn't imply that I will also love someone qualitatively identical. But, while this is so, it misses the point. If I rationally want just one piece of (a certain kind of) cheesecake, what I want is my having a single piece of cheesecake (of that kind). Which piece I have doesn't matter to me, if my desire is purely rational.

---

[15] At least, it disturbs me, and I think it would disturb Kath. It seems not to disturb Velleman in his (1999), p. 372. It does disturb Keller a little, but his discussion of it in his (2000), p. 171, is unsatisfactory. He seems simply to deny that two people will ever be sufficiently similar in the relevant way; this ignores the deeper question of whether I should, counterfactually, love Kay as I do Kath.

[16] Cf. Frankfurt (1999), p. 166: "Substituting some other object for the beloved is not an acceptable and perhaps not even an intelligible option. The significance to the lover of what he loves is not that of an exemplar; its importance to him is not generic, but ineluctably particular."

[17] Cf. Beardsley (1965), pp. 1-2 [* pp. 61-62 of this volume], for discussion of such a case.

Rationality doesn't require a special attachment to one piece rather than another; indeed, it requires that there be *no* such attachment. So too with Kath. If my love for her were rational, it might be based not just on her wit and figure but, more particularly, on being in the company of just one person with such a wit and figure; so that, while being in her company, I wouldn't welcome the company of a clone, should one walk by. But, again, in such a case it wouldn't matter who came first, Kath or the clone, whereas in fact this does matter very much to me.

In order to exploit the connection between value and the fittingness of pro-attitudes, the proponent of the view that individual objects can be the bearers of final value must find not only a pro-attitude that is directed toward individual objects, but one that is directed toward them for their own sakes and, moreover, one that there is a universalizable reason to direct toward them for their own sakes. I doubt that this can be done. Rabinowicz and Rønnow-Rasmussen might claim that this is exactly what they have done in the case of Diana's dress: its having belonged to Diana gives us good reason to treasure Diana's dress for its own sake, and this is duplicable in that Diana's shoes' having belonged to Diana gives us equally good reason to treasure them for their own sake. Here, though, I would once again point out that this just doesn't seem to be a case of treasuring the dress or the shoes nonderivatively, but rather to be a case where the value of the objects is being traced to some source (their having belonged to Diana or, more generally, their having a certain sort of connection to someone of a certain sort) which is itself the thing that is being treasured nonderivatively.

To this, though, Rabinowicz and Rønnow-Rasmussen might respond as follows.[18] Even if it were agreed that Diana's dress's having belonged to Diana is something that we value, it would be a mistake to say (as I have just done) that it is something that we treasure. But we do treasure the dress. Given that these attitudes are appropriate, this shows that the state and the object have different types of values, and thus that the latter's value cannot be derivative from the former after all.

This is tricky. I acknowledge that sometimes the attitude we have toward an object is different from the attitude we have toward some state of the object. (As I have noted, I have a romantic love for Kath but only a nonromantic love for her having the wit and figure that she has.) But it's not at all clear to me that we treasure only Diana's dress (on the assumption that we place any value on it at all) and not also its having belonged to Diana; on the contrary, I'm inclined to think that we do take the same attitude toward both object and state. Still, other cases could be cited. Rabinowicz and Rønnow-Rasmussen claim that we honor people but appreciate, rather than honor, their achievements, and that we respect people but appreciate, rather than respect, their courage.[19] Such cases could be multiplied. What do they show?

I don't think that they show that individual objects can have final value. First, it is again not at all clear to me that we do not in fact honor people's achievements or

[18] Cf. Rabinowicz and Rønnow-Rasmussen (2003).
[19] Rabinowicz and Rønnow-Rasmussen (2003).

respect their courage. Perhaps it is sometimes somewhat stilted to apply the same verb to both state and individual object, but that may just be an accident of language.[20] More importantly, even if it sometimes is the case that the type of attitude we have (and which it is appropriate to have) toward a state is distinct from the type of attitude we have (and which it is appropriate to have) toward the individual object involved, I cannot see that this shows that the object has final value. This would at best seem to follow only if the final value of something depended on the *whole* nature of the attitude that is appropriate to it, which in fact is dubious. If the whole nature of the attitude were relevant, then the fact that many instances of virtue are admirable but many instances of pleasure, though valuable, are not admirable would appear to imply that such instances of virtue are incomparable in value with such instances of pleasure. But there is no reason to think that this is so. As long as the different attitudes have a common "core," a common "denominator," this may be all that's strictly relevant to the determination of value and hence all that's needed for the states at issue to be comparable in value. And so, even if it is the case that Diana's dress is to be treasured but its having belonged to Diana is not, as long as there is a common core to the attitudes that are appropriate to both object and state, there is no reason to think that the former's value does not derive from the latter's.

However, suppose it were denied that this common core is all that's strictly relevant to the determination of certain values. Even then the conclusion that individual objects have, or can have, final value can be resisted. What exactly is it about the idea that individual objects have a different sort of value from the value that states have that requires us to say that the former value is final? Why could it not be, for instance, that, even if Diana's dress is to be treasured but its having belonged to Diana is to be valued in some other way, still the dress is to be treasured, not for its own sake, but for the sake of the state in question? That one entity derives its value from another would not appear to require that the two have exactly the same type of value. We should not be beguiled by talk of "source" and "derivation" into thinking that some single item is somehow being transferred from one locus to another. Indeed, despite the fact that it is customary to talk, as I have done and will continue to do, in terms of derivative *versus* nonderivative value, it seems to me that, when it comes to the relation between nonfinal and final value, it would probably be better to think of the former, not as being derivative from the latter, but rather as being *reflective* or *revelatory* of the latter. When something is

[20] Consider the following. Suppose that someone tells us that Diana's dress shocked the public. We ask why. The answer is that it was especially lavish. And then we say: "Ah, so it was the dress's being (so) lavish that shocked the public." Now suppose that someone tells us that Diana's suitcase broke the scale. We ask why. The answer is that it was especially heavy. To my ears, it would sound decidedly odd if we were then to say: "Ah, so it was the suitcase's being (so) heavy that broke the scale." May we infer from this that, although both objects and states may be properly said to be able to shock people, only objects and not states may be properly said to be able to break things? I doubt it. But even if we may infer this, such a fact about language would seem to provide no reason to deny that object-causation is parasitic on state-causation.

good not for its own sake but for the sake of something else to which it is in some way related, its value may be said to reflect or reveal the value in this something else. If we think in these terms, there is less temptation to suppose that nonfinal value must match final value in kind.

Fourth objection

If to admire John for his display of courage is no more than to admire his display of courage, and if to treasure Diana's dress for having belonged to Diana is no more than to treasure its having belonged to Diana, then one would expect us not only to rejoice in Peter's being pleased but to rejoice in Peter for being pleased. But we don't. This is an indication that to value Peter's being pleased is distinct from valuing Peter for being pleased, and thus that valuing John or Diana's dress is distinct from valuing the states involving them.[21]

Response

I concede the asymmetry. We admire John, we treasure Diana's dress, but we don't rejoice in Peter. I think this must have something to do with the particular pro-attitudes in question. I confess that I have no ready explanation either for why it is that we rejoice in Peter's being pleased but do not admire it or treasure it, or for why it is that the attitudes of admiration and treasuring can be directed toward both states and individual objects but the attitude of rejoicing-in apparently cannot. But even given the asymmetry, the objection is not persuasive. My claim is that it is John's display of courage that is worth admiring for its own sake, that it is Diana's dress having belonged to Diana that is worth treasuring for its own sake, and that it is Peter's being pleased that is worth rejoicing in for its own sake. If John can be admired and Diana's dress can be treasured but Peter cannot be rejoiced in, that does not alter the fact that it is the states that are worth valuing for their own sakes. If John is worth admiring, it is not for his own sake; if Diana's dress is worth treasuring, it is not for its own sake. In both cases the valuableness of the object is derivative from the valuableness of the relevant state. The fact that Peter isn't derivatively valuable doesn't somehow render the valuableness of John and of Diana's dress nonderivative. The thesis that, when an object has value, this value is parasitic on the value of some state, isn't undermined by the observation that some types of value don't sanction derivation of value from state to object.

[21] See Rabinowicz and Rønnow-Rasmussen (2003).

Fifth objection

Even if the value of individual objects such as Diana's dress is derivative from the value of states such as the dress's having belonged to Diana, such an account simply cannot be accepted when the individual objects in question are persons. Kant has taught us that persons are uniquely valuable; they have a "dignity" that requires that they be "exalted above all price."[22] It would be a gross distortion of his view to say that it is not persons that have such value but states involving persons that do. This would simply misidentify what Kant takes to be the locus of value.

Response

Whether persons are indeed to be "exalted above all price" is a very difficult issue that I won't try to resolve here. The idea that they are to be so regarded is a powerful and welcome repudiation of the excesses of classical utilitarianism, which licenses the mistreatment of individual persons for the good of the many. But of course the Kantian view of persons faces difficulties of its own. There is no need to rehearse these difficulties here, however; for, even if Kant is right, this poses no problem for my view that it is states that are the bearers of final value.

The fact is that the sort of value that Kant attributes to persons is not the sort of value that I have called final value and have attributed to states. One indication of this is that, in saying that persons are to be "exalted above all price," Kant is evidently claiming that all persons have infinite worth. If this were understood as a thesis about final value, it would present at least two difficulties.

First, it would preclude the sorts of meaningful comparisons regarding final value that it is natural to make. Suppose, contrary to what I have said above, that John does have final value in virtue of his display of courage, and that Kath also has final value in virtue of her wit and figure. Wouldn't it be an extraordinary coincidence that they have precisely the same value? Suppose that Mary has none of the virtues of John and Kath and no other redeeming features, either; suppose, indeed, that all her personal qualities are reprehensible. Wouldn't it be absurd to say that she has a final value as great as John's and Kath's?[23]

Second, if the value that Kant attributes to persons were final value, it would seem to follow that all worlds in which a person exists are infinitely good (unless Kant were to invoke some bizarre instance of the principle of organic unities according to which a "whole" that contains persons can somehow be worse than its "parts"). It is simply not credible that all such worlds are infinitely good, and I don't for a moment think that Kant would maintain that they were. In his discussion of

[22] Kant (1964), p. 102 (Ak. 434-5).

[23] *Pace* Kant, people can surely differ in value in ethically relevant ways. For example, some people are morally better than others, and this is a value that attaches directly to them. In my opinion, this is not a matter of their having greater final value, however. Cf. Lemos (1994), p. 27 [* p. 187 of this volume].

dignity, Kant just isn't concerned with final value. He's concerned with the normative issue of how we are to treat persons, not the axiological issue of how good persons are.[24]

Sixth objection

Even if it is true that Diana's dress is valuable *only if* some state involving it is valuable, this doesn't show that the value of the dress is derivative from the value of the state. For it is equally plausible to say that the dress is valuable *if* the state is. We thus have an equivalence. No reason has been given to prefer reduction in one direction (from object to state) to reduction in the other (from state to object).

Response

Such reason has been given. Just as one attains a fuller understanding of what transpired once it is revealed that it is in particular the dress's being lavish that caused the sensation, so too one attains a fuller understanding of what is at issue once it is revealed that it is in particular the dress's having belonged to Diana that is valuable. This indicates that the state is more fundamental than the object, relative to the context in question.

There is a further important advantage to the view that I am advocating here. Almost everyone seems to agree that states (such as Peter's being pleased) can have final value.[25] If it is also admitted that *only* states can have final value, matters are simplified considerably. In particular, one can dare hope that an informative and helpful account of the *computation* of final value might eventually be provided. It is very hard to see how any such account could be given if the bearers of value were ontologically mixed. The prospects are much brighter if the bearers are ontologically uniform.[26]

Of course, even if the bearers of value are ontologically uniform, a useful account of the computation of value might still prove elusive. There is the possibility that the principle of organic unities is true, and this would pose a problem for any account of computation. There is also the fact that different attitudes seem appropriate to different states. As noted earlier, John's display of courage seems admirable whereas Peter's pleasure does not. This suggests that the values of these

---

[24] This interpretation is confirmed by the following passage in Kant (1997), p. 44:

> What does the highest good consist in? The most perfect world is the highest created good. But the most perfect world involves the happiness of rational creatures and the worthiness of these creatures for such happiness… If the world were full of…rational creatures, who were all well behaved, and thus worthy of happiness, and they were in the neediest circumstances, surrounded with sorrow and trouble, they would then have no happiness, and there would thus be no highest good there.

[25] Anderson may be an exception. See her (1993), pp. 20 and 26.

[26] Note that, just as one must guard against overcounting causes, for fear of diagnosing overdetermination where there is none, so too one must guard against overcounting value, for fear of mistaking the world for being better or worse than it really is.

states may be incommensurable. I believe that both these problems are soluble, but this is not the place to address them.

4.

If states are the bearers of final value, then the traditional talk of "intrinsic value" rather than "final value" is innocuous. For the final value of states supervenes on an intrinsic property that they have, namely, the property of having a certain constituent property. For example, the final value of Diana's dress having belonged to Diana supervenes on its having the property of having belonged to Diana as a constituent. This is part of the very nature of the state; as such, it is not only intrinsic to the state but essential to it.

Ingmar Persson has told me, in his distinctive manner, that he finds it rather peculiar to use the term "intrinsic value" in this way. In his view, even if states are the bearers of final value, the term "intrinsic value" should be restricted to those cases where the constituent property of the state is intrinsic to *its* bearer. Thus Peter's being pleased may be said to be not just finally but intrinsically good, since the property of being pleased is intrinsic to Peter. But Diana's dress having belonged to Diana may not be said to be intrinsically good, even if it is finally good, since the property of having belonged to Diana is not intrinsic to the dress.

This is not a matter of great moment. The dispute is merely terminological. However, I continue to think that "intrinsic value" is properly used when used in the way that I have proposed. This is because it seems to me most natural to say that something has intrinsic value just in case the value in question supervenes on and only on one or more of *its* intrinsic properties. This is precisely the case with the final value of states.[27]

BIBLIOGRAPHY

Anderson, Elizabeth (1993). *Value in Ethics and Economics*. Cambridge, Mass.: Harvard University Press.

Beardsley, Monroe C. (1965). "Intrinsic Value." *Philosophy and Phenomenological Research*, 26: 1-17 [* pp. 61-75 of this volume].

Brentano, Franz (1969). *The Origin of Our Knowledge of Right and Wrong*. London: Routledge and Kegan Paul.

Broad, C. D. (1930). *Five Types of Ethical Theory*. London: Kegan Paul, Trench, Trubner.

Ewing, A. C. (1948). *The Definition of Good*. New York: Macmillan and Co.

Frankfurt, Harry G. (1999). *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

Kagan, Shelly (1998). "Rethinking Intrinsic Value." *Journal of Ethics*, 2: 277-97 [* pp. 97-114 of this volume].

[27] I have enjoyed and profited from discussions with Ingmar Persson about the issues treated in this paper. Many thanks also to Ben Bradley, Krister Bykvist, Erik Carlson, Toni Rønnow-Rasmussen, and especially Wlodek Rabinowicz for their help.

Kant, Immanuel (1964). *Groundwork of the Metaphysic of Morals*. New York: Harper and Row.
Kant, Immanuel (1997). *Lectures on Ethics*. Cambridge: Cambridge University Press.
Keller, Simon (2000). "How Do I Love Thee? Let Me Count the Properties." *American Philosophical Quarterly*, 37: 163-73.
Korsgaard, Christine M. (1983). "Two Distinctions in Goodness." *Philosophical Review*, 92: 169-95 [* pp. 77-96 of this volume].
Korsgaard, Christine M. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
Lemos, Noah M. (1994). *Intrinsic Value*. Cambridge: Cambridge University Press.
Moore, G. E. (1912). *Ethics*. Oxford: Oxford University Press.
Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (1999). "A Distinction in Value: Intrinsic and For Its Own Sake." *Proceedings of the Aristotelian Society*, 100: 33-52 [* pp. 115-29 of this volume].
Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (2003). "Tropic of Value." *Philosophy and Phenomenological Research*, 66: 389-403 [* pp. 213-26 of this volume].
Ross, W. D. (1939). *Foundations of Ethics*. Oxford: Oxford University Press.
Rønnow-Rasmussen, Toni (2002). "Instrumental Values – Strong and Weak." *Ethical Theory and Moral Practice*, 5: 23-43.
Velleman, J. David (1999). "Love as a Moral Emotion." *Ethics*, 109: 338-74.

CHAPTER 16

T. TÄNNSJÖ

A CONCRETE VIEW OF INTRINSIC VALUE

1. INTRODUCTION

There are no genuine problems of value. The basic questions of ethics are different. They are the questions of which actions are right and wrong, and in virtue of what are right actions right and wrong actions wrong. In the final analysis, we want an answer to these questions that is put in empirical terms. We want a criterion of rightness cast in empirical terms. If we succeed in formulating such a criterion, then we do not need the notion of intrinsic value. There is no fundamental moral problem, in addition to the normative problem, about what is good in itself. However, the notion of value has an epistemic role to play in our moral thinking. When we say that something has intrinsic value, we say that it is something we have to consider in our normative judgments, even though we are not sure how exactly we should take it into account. We may feel, for example, that happiness has some kind of moral significance, even though we are not sure whether it should be maximized in the world or not. We may put this as follows: happiness is a good-making characteristic, states of happiness are good in themselves.

If this is correct it means that the notion of intrinsic value is a kind of blank that should eventually be replaced by an empirical specification. Yet, for all that, it might be interesting to speculate about what it is that has intrinsic value. People who hold on to different normative principles may agree about intrinsic value. But before we can speculate about what it is that possesses intrinsic value, we should be more precise about the ontological status of the subject matter.

2. CONCRETISM VERSUS ABSTRACTISM

What may possess intrinsic value? G.E. Moore spoke about *things* having intrinsic value. That is on the right track, though most people discussing the subject today tend to think otherwise. They argue that the things that possess intrinsic value are things such as facts, propositions, or states of affairs.

Facts, states of affairs, and propositions are abstract entities. They may be more or less specific. The state of affairs consisting in my being pleased now is, for example, more specific than the state of affairs consisting in someone being satisfied

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 207-211.
© 2005 *Springer. Printed in the Netherlands.*

sometimes. However, even a very specific state of affairs is abstract in the sense that it does not form a part of the concrete world. It does not occupy a place in time and space. We may instead conceive of it as a very complex proposition.

Material things do occupy a place in time and space. However, it is not satisfactory to say of material things that they are good or bad in themselves. Instead, I think we should say that what are good in themselves are concrete processes in space and time such as the state of consciousness I am actually in when I write this.

While it is natural to identify a fact with a true proposition, and a true proposition with a state of affairs that obtains, many propositions may give true accounts of one and the same concrete process in the world. It may be true of my mental process right now that it is a state of happiness, that it is a state of happiness entertained by a philosopher writing a paper on intrinsic value, and that it is a state of consciousness with a more intense hedonistic tone than the state the same person was in before he began to write the article. All these are different facts, but there is no corresponding multitude of concrete mental states. I am in the state I am in, and that is it.

How do I identify the mental state I am in? The temporal borders could be very exactly specified. The spatial delineation will be given with reference to my body, or the parts of my body responsible for mental characteristics of it, during the time interval in question. Might not two concrete states or processes exist in the same spatial and temporal location? An example of this could be a sphere spinning on two axes at once.[1] But we should resist the temptation to answer this question in the affirmative. It is preferable to say of one concrete existing process, specified by its temporal and spatial coordinates, that it involves two movements of the sphere.

3. INTRINSIC VALUE

On a concretist understanding of intrinsic value, what does it mean to say of something valuable that it has intrinsic value? To answer this, we may use G. E. Moore's isolation test. After posing the question "What things have intrinsic value, and in what degrees?" Moore goes on to say:

> In order to arrive at a correct decision on the first part of the question, it is necessary to consider what things are such that, if they existed *by themselves*, in absolute isolation, we should yet judge their existence to be good; and, in order to decide upon the relative *degree* of value of different things, we must similarly consider what comparative value seems to attach to the isolated existence of each.[2]

What is the intrinsic value of the state of consciousness I am in right now? On Moore's test, its intrinsic value is equal to the value of a world consisting of the

---

[1] This objection has been raised by Michael J. Zimmerman in correspondence.

[2] Cf. *Principia Ethica* (Cambridge, England: Cambridge University Press, 1965), p. 187.

mental process and nothing else. We can conceive of such a world because we conceive of it concretely. In such a world only a certain time-slice of me exists.

A fact cannot exist in isolation, however. Therefore the isolation test is not open to those who conceive of intrinsic value as abstract. Not even an exhaustive description of my state of affairs would be an example of an abstract fact existing in isolation. First of all, to isolate this fact, we would have to add to the exhaustive description that it is exhaustive. Furthermore, along with the exhaustive description go all kinds of disjunctive states of affairs, and negative states of affairs. What are their value? An extensive literature on such questions exists, starting with Roderick Chisholm, but no suggested answer seems to be satisfying.[3] The questions tend to boggle the mind. It is therefore a merit of the concretist view that we need no answer to them. It is also a merit, of course, that the isolation test for intrinsic value can be used.

4. INTRINSIC VALUE AND GOOD-MAKING CHARACTERISTICS

It is true that sometimes we say such things as that pleasure and knowledge have intrinsic value. But pleasure and knowledge do not seem to be concrete entities. We should understand such statements as conjectures about what it is that makes concrete intrinsically valuable processes valuable.

Suppose my state of affairs now can be considered to be of positive intrinsic value of a certain magnitude. We consider other states and find that they have the same degree of intrinsic value, the value that would exist in the world if they existed alone. Suppose we find that something the states have in common is that they are pleasurable to a certain degree. We may then conjecture that pleasure is a good-making characteristic and that intrinsic value is a function of the intensity of pleasure felt. But suppose we come across an example of undeserved pleasure, which we find is of negative value. If we want to stick to our basic intuitions about such cases, we have to revise our hypothesis. We have to withdraw our initial conjecture that happiness is a good-making feature. We have to say something like: while deserved happiness is a good-making characteristic, undeserved happiness is not and may even be a bad-making characteristic.

There is much to be added about how we form responses to concrete states of the world, conceived in isolation, but, in principle, the method by which we gain evaluative knowledge is a standard inductive procedure. We make conjectures and

[3] "The Intrinsic Value in Disjunctive States of Affairs", *Noûs*, 9 (1975), pp. 295-308 [* pp. 229-39 of this volume]. Recent contributions to this discussion include Erik Carlson, "The Intrinsic Value of Non-Basic States of Affairs," *Philosophical Studies*, 85 (1997), pp. 95-107 [* pp. 361-70 of this volume]; Sven Danielsson, "Harman's Equation and the Additivity of Intrinsic Value," in Lindahl, Needham and Sliwinski, eds., *For Good Measure: Philosophical Essays dedicated to Jan Odelstad on the Occasion of his Fiftieth Birthday* (*Uppsala Philosophical Studies*, 46 (1997), pp. 23-34); and Thomas Magnell, "Evaluations as Assessments, Part I: Properties and Their Signifiers," *Journal of Value Inquiry*, 27 (1993), pp. 1-11, and "Evaluations as Assessments, Part II: Distinguishing Assertions and Instancing Good of a Kind," *Journal of Value Inquiry*, 27 (1993), pp. 152-63.

we test them in concrete cases. Or, more correctly, we test them in relation to thought experiments, where we abstract from everything besides a certain concrete process which we consider as if it existed alone in the universe. In many cases, since the processes we focus on are our own mental processes, we may claim to have a direct knowledge of them.

5. ORGANIC UNITIES

Moore is known, not only for his isolation test, but also for his conjecture that there may exist organic wholes, or compounds of intrinsically valuable things that are more or less valuable than the sum of the values of their parts. Concretism makes this view comprehensible.

Consider an extremely simple case. Suppose we find that a certain happy state of consciousness has the value +5 and another happy state of consciousness has the value +100. On the isolation test, this means that these are the values two worlds would have, had each of the states existed in isolation. What of the intrinsic value of the two states existing together, however? There is nothing mysterious in the talk of the two states existing together. When we want to find out about the value of the two discrete states existing together, we concatenate them by conceiving of a world where they, and no other concrete states, exist, and we assess the value of the world by associating it with a real number. All this is in accordance with standard theory of measurement. This is very different and much more simple than the corresponding abstract idea of concatenating facts by combining them with the logical operation of conjunction.

Suppose we find that the value of the world where both these states of happiness exist together is +50. Perhaps nothing like this will actually happen, when we assess the value of a world. But it is not impossible or incomprehensible that it should happen. If it were to happen, we would have to say that Moore was right in his conjecture that there may exist organic wholes.

We would want an explanation of the fact that the two concrete processes, when combined in one world, have less value than the sum of the intrinsic values each of them possesses. One possibility could be that the distribution of pleasure is of importance as well. The unequal distribution in the world under consideration may be a bad-making characteristic of it.

If we came across such an organic whole, this would not mean that we would have to withdraw our conjecture that pleasure is a good-making characteristic. It might still be true that it is the fact that it is pleasant which makes a concrete experience intrinsically good. It might also be true that, were I to have a concrete and deservedly pleasant experience, that experience would be good. Yet, the experience might form part of a whole possessing a value different from the sum of values of its parts. If there are organic wholes, they do threaten to make a mess out of our moral mathematics.

A hard-nosed hedonistic utilitarian may deny that organic unities exist.[4] However, on the concretist view, it is an open question whether they exist or not, and this is a further merit of the view. The concretist view makes sense of Moore's idea that there are organic wholes.

6. CONCLUSION

Moore used to think concretely of intrinsic value. He thought that things were good or bad in themselves. The test by which it could be ascertained whether a thing was good or bad was the isolation test. Consider the thing in question in isolation. Consider a world where it, and no other object, exists, and assess the value of this world. The value of this world is the intrinsic value of the thing in question.

Moore was on the right track. It was wise of him to consider concrete entities as bearers of intrinsic value. However, it is more natural to take concrete processes or events to be bearers of intrinsic value than to take things to be such bearers. Concrete processes or events can be put to the isolation test just as well as things can be put to the test. This is a merit in the concretist view, since it is hard to see how an abstract state of affairs could be tested in any similar manner.

It is simple too to conceive of measurement of intrinsic value, if the bearers of intrinsic value are taken to be concretely existing processes. We measure the intrinsic value of concrete processes in the way we measure, say, the length of persons. The operation of concatenation of processes is no more mysterious than the operation of concatenation of persons, and we assign to discrete concrete processes real numbers when measuring intrinsic value in the way we do with persons, when we measure their length.

This is not to say that there are no unsolved problems of measurement in relation to intrinsic value. Is there a natural zero, for example? Is the value of two processes, that we have concatenated by bringing them into existence in the world, equal to the sum of the respective values of the worlds containing each of them? If we concatenate something with itself, do we achieve twice its value then? This question cannot be answered before we have settled on an answer to the question: "What is it that has intrinsic value?" The more complicated the account, the more theoretical and practical problems of measurement are likely to surface.[5]

---

[4] I defend this view in my recent book, *Hedonistic Utilitarianism* (Edinburgh: Edinburgh University Press, 1998).

[5] I thank Lars Bergström, Sven Danielsson, Thomas Magnell, Hans Mathlein and Michael J. Zimmerman for valuable comments.

CHAPTER 17

W. RABINOWICZ AND T. RØNNOW-RASMUSSEN

TROPIC OF VALUE

1. INTRODUCTION

In Rabinowicz & Rønnow-Rasmussen (1999), we defended the view that not only abstract states of affairs but also concrete objects, such as things and persons, may have final value, i.e., value for their own sake.[1] If this view is accepted, then it becomes possible to argue that a final value need not be intrinsic, i.e., that it need not be exclusively dependent on the internal (= non-relational) properties of its bearer. Thus, to give an example, a dress of Princess Diana may well be seen by some as valuable for its own sake in virtue of its relation to Diana. Likewise, if a tropical wilderness is ascribed a value for its own sake, then this value may be enhanced if the wilderness has never been visited by humans. Or, to give an example not mentioned in our previous paper, if knowledge (true well-grounded belief) is taken to be valuable for its own sake, then one must allow that a belief of a person can acquire final value partly because of its relation to its object. On the externalist analyses of knowledge, an appropriate relation to an object is required not only to make a belief true, but also to make it well-grounded. Examples like this can be easily multiplied. (For a somewhat different list, cf. Kagan 1998). But they all seem to depend on allowing concrete entities as potential bearers of final value. If all such value were instead taken to belong to facts or states of affairs, i.e., to abstract entities of a "propositional" kind, then it would be possible to maintain that all final value is intrinsic. For then one might import the relevant relational property of a concrete object into a state of affairs that deals with the object in question. If it is some such state rather than the object itself that should be ascribed final value, then one might insist that the final value of a state of affairs only depends on its internal features, such as its having certain intrinsic constituents or components.

Therefore, in our 1999 paper, we have considered how one might argue for the view that the final value of things or persons can be reduced to the final value of states. The particular form of reduction we have discussed makes the following claim: whenever a concrete object, say, a thing or a person, is said to have final

[1] A caveat: this claim is about final value, not about final goodness. As the term "good" is ordinarily used, it is inappropriate to say of a thing or a person *a* that *a* is *good* for its own sake. It is different with ascriptions of *value*.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 213-226.
© 2005 *Springer. Printed in the Netherlands.*

value in virtue of its property P, which may or may not be relational, then what "really" has final value is that this object, which has P, exists. Or, perhaps, even more simply, what has final value is that there exists something that has P. Thus, when a dress is said to be finally valuable because of its special relation to Diana, the reduction manoeuvre under consideration locates the final value in the state consisting in the existence of a dress that has this special relation to Diana.

2. OBJECTIONS TO THIS REDUCTION

In the paper mentioned above, we present several objections to this reduction proposal. One of them is that reduction of this type puts the cart before the horse: why do we think that it is of value that Diana's dress exists if not because we take the dress itself to be valuable? Or, why is the existence of the wilderness valuable if not because of the value of the wilderness itself? The value of the thing thus seems to be ontologically prior to the value of the state: the former grounds the latter and not vice versa.[2] Another objection might be that ascribing value to an object, such as Diana's dress, need not *ipso facto* commit us to ascribing value to its existence. For some valuable objects (such as, say, the wilderness in our example), their existence may be seen as valuable, but not necessarily for all.

Apart from the objections to that particular reduction proposal, there is also a line of criticism that applies to all the views according to which states alone can be the bearers of value. Consider the following assumptions about the concept of final value, all of which seem to be plausible:

(i) To be valuable, for its own sake, is to be a fitting object of a positive response (a pro-attitude or a pro-behaviour) that is directed to the value bearer for its own sake.[3]

(ii) Such positive responses can be of different kinds (desiring, promoting, caring for, treasuring, rejoicing in, admiring, loving).

---

[2] This talk of ontological priority relations among values should not be taken to suggest a commitment to value objectivism. Even subjectivist conceptions of value can leave room for quasi-ontological relationships between the values we project on the world. For a further discussion of some issues involved in giving such a projectivist account of final value, see our (1999) [* pp. 115-29 of this volume].

Note that we do not claim that a valuable state of the form *there exists some object that has P*, or *this object, which has P, exists* must *always* draw its value from the value of the object involved. It is even conceivable that there might be cases in which the thing itself lacks value even though its existence is finally valuable. However, in the examples considered above, it is not like this. In these cases, it does seem natural to say that the existence of an object is valuable because of the value of the object itself.

[3] This idea goes back at least to Brentano (1969 [1889]). Cf. also Broad (1930), Ross (1939), Ewing (1947), Lemos (1994) [* pp. 17-31 of this volume]. Note that negative value can be dealt with in an analogous way: to be negatively valuable is to call for a negative response. Note also that, if there is to be a conceptual link between the fitting response towards an object and that object's value, the fittingness of the response must be assumed to be independent of any pragmatic considerations (such as, for example, that having responses like this towards objects like that would make us happier or have some other good effects for us or for other people).

(iii) The kinds of positive responses that are appropriate with respect to a valuable object need not be appropriate, and may even be impossible, with respect to another valuable object. Thus, to illustrate, you can admire a person but you cannot admire a state of affairs; you can rejoice in a state, but you cannot rejoice in a person.[4] Or, to take another example, some things or persons worthy of, say, protection or treasuring need not be worthy of admiration.

From these assumptions it follows, first, that final values may be quite heterogeneous, depending on the kinds of positive responses the valuable objects call for.[5] Secondly, since some of the relevant responses may be thing- or person-oriented, it follows that not only states but also things or persons may have final value. It would appear that the ubiquitous reductive tendencies towards thing- and person-values may be grounded in a philosophical ambition to reduce all responses that are fitting with respect to valuable objects to just *one* kind, such as preferring or promoting. Since preferring as well as promoting take states as their objects, it becomes natural from this point of view to reduce all value to the value of states. However, as soon as a larger range of positive responses is allowed for, the reductivist tendency loses its appeal.

3. FINALITY, NON-DERIVABILITY, AND NORMATIVE RELEVANCE

This general argument for ascribing final value to things or persons may be resisted. One might insist that in all those cases in which it is fitting to hold a pro-attitude (or to engage in a pro-behaviour) towards a thing or a person *a*, this must depend on *a* being a constituent in *some states or others* that themselves are fitting objects of pro-attitudes (or pro-behaviours). While the value of the state consisting in the *existence* of a valuable thing seems to depend on the value of that thing, the latter value might in its turn depend on the value of some other states of affairs. If this were true, then the value of *a* would be derivative from the value of the states in question. Now, it might be thought that what makes a value final is that this value is non-derivative from other values. On this assumption, the putative dependence of thing- and person-values on the values of some states would imply that the former values are not final. However, on our view, the claim that a certain value is final does *not* imply that it must be non-derivative. We interpret final value, i.e., value for its own sake, in the standard way, by contrasting it with value as a means (instrumental value) and value as a part (contributive value). What is non-derivatively valuable must be valuable for its own sake, but not necessarily vice versa. Thus, for example,

---

[4] But can't we admire, say, Socrates' resolute determination to abide by the laws of his country? Yes, we can, but the admiration applies, in this case, to Socrates himself or perhaps to his character, rather than to some state of affairs.

[5] This value pluralism based on the plurality of fitting responses is perhaps most clearly developed in Baron (1997). Cf. also Anderson (1993) and Swanton (1995).

a whole that consists of several parts may be valuable for its own sake and still this value may be derivative from the value of its parts: The whole may be valuable because of the valuable parts that it contains (cf. Brülde 1998, pp. 7 and 392). Many other examples of derivative final values can be given, some of them deriving from our own discussion. For example, one can value, for its own sake, the state of affairs consisting in the existence of a wilderness that has never been touched by humans. But one can still maintain that the final value of this state is derivative, insofar as it depends on the wilderness itself having value.[6]

For our concerns, the notion of derived value is relevant in only one way: if the value of X is in some sense derived from the value of Y, then the latter value is not reducible to the former. That was the point of the "cart before the horse"-objection. But nothing hinders that both values can be final. Nor is derivability the same thing as reducibility: while reduction is a form of derivation, we can also have derivation without reducibility. The value of X may be derived from, or grounded in, the value of Y without being reducible to it. Thus, while the value of a state consisting in the existence of a certain concrete object appears to be derivable from the value of that object itself, the two values should still be distinguished. Insofar as the fitting responses towards the state (preferring, promoting, etc.) differ from the fitting responses towards the object (protecting, admiring, etc.), it would be misleading to say that the value of the former is "nothing but" the value of the latter.

It might be argued, however, that final values of things and persons, even if they are non-reducible, still lack interest from a *normative* point of view. Only the final values of states have normative relevance. Or, at least, so it might seem from a consequentialist point of view, according to which the normative status of actions exclusively depends on the value of states realised by these actions and by their alternatives. Thus, from a normative perspective, the final values of states are the only ones that deserve attention.[7]

We believe that this objection is not justified, not only because the consequentialist view of morality may well be contested but also, most importantly, because even on that view the values of some states could be seen as grounded in the values of things or persons. Thus, the latter values would have at least an indirect normative relevance. Furthermore, what is at issue from the normative point of view is not only what actions we should perform but also what attitudes we should take towards various objects, including things and persons. For this latter question, the final values of these objects are directly normatively relevant, given the existence of a conceptual link between final value and fitting pro-attitudes.

---

[6] What such dependence of one value on another is supposed to consist in is a difficult issue, which we prefer not to address in this paper. To clarify what makes a value derivative is a task that deserves a separate investigation.

[7] We are indebted to Krister Bykvist for raising this issue.

4. A NEW REDUCTIVIST PROPOSAL

On our view, final values of things and persons are neither reducible nor normatively irrelevant. Still, the matter is complex and we may be wrong. Therefore, we will consider an alternative reduction proposal with respect to such values that has been put forward by several commentators, in particular by Ingmar Persson and Michael Zimmerman. Persson was the first to suggest it to us, in discussion, and Zimmerman recently defended it in his (2001b). In our 1999-paper, we briefly discuss that proposal in a footnote, but we now realise that it should be given more serious attention.

Suppose a thing (or a person) *a* is said to be valuable for its own sake. Its putative final value must then supervene on some of its properties. For simplicity, assume that P is the conjunction of all the evaluatively relevant properties of *a*. Thus, P is a possibly very complex property of *a* on which *a*'s final value is supposed to supervene. On the reduction proposal under consideration, the final value is again located in a certain state of affairs, but the relevant state, rather than being existential in form, simply consists in that *a has P*. For example, in the case of *a* being Diana's dress, the state that bears the final value might be the one that consists in that *a is a dress that has belonged to Diana*.[8] Similarly, in the case of *a* being the tropical wilderness, the final value might accrue to the state that *a is a tropical wilderness that has never been touched by humans*.[9]

What is it about this proposal that makes it preferable to the type of reduction that we already have rejected? As may be recalled, one of the objections against that reduction was that it put the cart before the horse. The final value of Diana's dress is not reducible to the final value of the existence of that dress, simply because the value of the former appears to be ontologically prior to the value of the latter, not the other way round. Now, as Ingmar Persson was the first to note, there is no similar objection to the present reduction proposal. It would be counter-intuitive to suggest that the value of *a* is ontologically prior to the value of *a has P*. Zimmerman makes the same point. In fact, he goes further and suggests that the order of derivation is the opposite one: the value of Diana's dress is properly accountable for by the value of that dress having belonged to Diana.

---

[8] Actually, Zimmerman would instead want to locate final value in some state that is more "fundamental" from the axiological point of view, such as, say, that *a bears an intimate causal relation to an important historical personality*, for example. This has to do with Zimmerman's identification of final value with non-derivative value. In what follows, we shall ignore this complication.

[9] More precisely, Zimmerman suggests that the proper value bearers are not states of affairs but *facts*, where the latter may be seen as the obtainings of states of affairs. A state of affairs may exist without obtaining, just as a property may exist as a universal without being instantiated. We can refer to a state such as, say, John being happy even when John in fact is miserable. For facts, on the other hand, this distinction between existence and obtaining cannot be made. For a discussion of the idea of facts as value bearers, see Appendix below. The distinction between states and facts, while important, does not make much difference as far as our discussion is concerned. Our objections to the reduction of thing- and person-values to the value of states apply equally well if states are replaced with facts.

We agree that the "cart before the horse"-objection is not applicable to the reduction proposal under consideration. For this reason, we think that the proposal in question is worth serious attention. However, we are inclined to reject it. It is true, by hypothesis, that the value of an object *a* is accountable for by *a* having P. But we are not prepared to agree with Zimmerman's suggestion that the value of *a* is accountable for by the *value* of *a* having P. Rather, it would be more natural to say that the value of *a* can be accounted for by (i) P being a value-making property (i.e., a property that makes its bearer valuable), together with the fact (ii) that *a* has P.

Indeed, it may well be doubted whether a state such as *a has P* is valuable. In a harmless sense, we can say of course that it is valuable that *a* has P, meaning by this no more than that *a* is valuable because *a* has P. But if *a has P* is claimed to be valuable on its own, so to speak, this gives rise to a puzzle: why should it be valuable that this particular dress has belonged to Diana? Certainly, one might say, its having belonged to Diana makes *it* valuable, but why should it be of any value that this dress, rather than some other one, has belonged to Diana? To put this worry concerning the value of *a* having P in a more general way, it is unclear what contribution the object *a* is making to the value of that state. Is it important for the putative value of the state in question that it consists in *a*, rather than some other object, having P? Surely not! So what is *a* doing there?

But then, as someone might suggest, perhaps what has value is not so much that *a* has P but rather that *something (some object or other) has P*? Such a response, however, would be unsatisfactory. To state that something has P is just to state that there exists an object that has P. And, according to the "cart before the horse"-objection, the value of the latter state comes from the value of the object itself. Furthermore, the latter state would obtain as soon as there were at least one object that instantiated P: increasing the number of such objects would add nothing to the truth of the existential statement. But we would want to allow for the possibility that increasing the number of such objects does increase the amount of value in the world, at least in some cases. (For this point, we are indebted to Sten Lindström.) It is good that there is a wilderness untouched by human hands, but it would be even better if there were several areas like this.[10]

5. TROPES AS VALUE BEARERS

Perhaps, then, the answer to our worry is that the role of *a* is to make it clear that what is of value is each particular instantiation of P rather than the mere fact that P is instantiated by some object or other. That the instantiation of P occurs in *a* rather than, say, in another object *b*, does not matter for our valuation, however. What is valued is not that *this* object *a* has P but rather the instantiation of P, which happens to occur in *a*. We value, in the same way, each instantiation of P, in whatever object

[10] However, if there are lots of such areas, the increase in their number may not be desirable. Somewhat similarly, if there were very many dresses that have belonged to Diana, the discovery of one more dress like this might not be welcome.

it occurs. We value, then, various instantiations of P rather than the states that *a* has P, that *b* has P, etc. Unlike the latter, the instantiations of P in various objects do not contain these objects themselves as constituents.

But can we differentiate between the *state* that consists in *a* having P and the *instantiation* of P by *a*? Is there any difference between *a* having P (a state) and *a*'s having P (an instantiation)? Aren't these the same? Not really. The former is a state of affairs while the latter is a state of an object. Even though the state of affairs that consists in *a* having P obtains if and only if *a* instantiates P, this instantiation of a property by an object may be seen as an entity *sui generis*, which should be distinguished from the associated state of affairs. Among ontologists, such entities have come to be known as *tropes*. If the instantiation of P in *a*, which we might refer to as *a*'s having P, is an entity *sui generis* that does not contain *a* itself as a constituent, this would make sense of the suggestion above that it is not *a* having P (a state) but rather *a*'s having P (a trope) that has value.

The term "trope" was put to use in ontology by Donald Williams (1966, 1986 and 1997 [1953]). Trope ontology was then further explored by Keith Campbell (1997 [1981], and 1990) and by John Bacon (1995, 1997), among others. Tropes are often characterised as *abstract particulars*. Thus, Campbell writes:

> The colour of this pea, the temperature of that wire, the solidity of this bell, are *abstract* in this sense only: that they (ordinarily) occur in conjunction with many other instances of qualities (all the other features of the pea, this piece of wire or the bell), and that, therefore, they can be *brought before the mind* only by a process of selection, of systematic setting aside, of these other qualities of which we are aware. Such a selective process is an act of abstraction… *Abstract* here contrasts with *concrete*: a concrete entity [the pea, the piece of wire, the bell] is the totality of the being to be found where our colours, or temperatures or solidities are… And our abstract particulars are particulars because they have a local habitation… They exist as individuals at unique place-times. The case of colour which belongs to this pea is the colour of this pea and no other… it cannot be instantiated in many different situations simultaneously. They are particular in just the same way and for the same reason that the pea, or the wire or the bell are particular. (Campbell (1990), pp. 2 ff; cf. Williams (1966), p. 78, and (1997), p. 113.)

We have characterized tropes as instantiations of properties. One and the same property (universal) may have several instantiations (exemplifications), in different objects. But this characterization of tropes assumes that properties (universals) exist, in the first place. Most tropists would deny this; they would refuse to admit universalia into their ontology. From that perspective, tropes come as a radical solution to the age-old difficulties about universals: they replace universals rather than exemplify them. We might then continue to speak about tropes as "instantiations of properties", for reasons of convenience, but we should be careful not to take this way of speaking literally. On this radical tropist view, it is tropes that *are* properties. What corresponds to universals in this ontology are maximal classes of tropes that are pairwise related to each other by the relation of similarity. Some

tropists are prepared to entertain just two kinds of *basic* entities: things ("substances") and tropes, while others go even further and treat things as (maximal) collections or clusters of tropes that are pairwise related to each other by the relation of "compresence".

In this paper, we avoid taking a stand on such fundamental ontological issues; in fact, we leave it open whether a satisfactory ontology can avoid tropes altogether. It is fair to say that the adherents of tropes form a very small minority among ontologists.[11] Still, we would like to consider a conditional question: *if* tropes exist, can they be value bearers and, in particular, can the final values of things or persons be reduced to the final values of tropes?[12]

As for the question whether tropes can be value bearers, Williams suggests that they are particularly well fit for this role:

> Evaluation is … focussed on abstracta. What most men value the moon for is its brightness; what a child wants of a lollipop is a certain flavor and endurance. He would much rather have these without the rest of the bulk than the bulk without the qualities. (Williams, 1997 [1953], p. 123)

Keith Campbell continues on the same track and suggests that tropes help us to deal with problems that arise in connection with conflicting evaluations of one and the same concrete object:

> Evaluation is another field in which the admission of tropes does away with awkwardness. Concrete particulars can be simultaneously subject to conflicting evaluations – in different respects, of course. A wine's flavor can be admirable and its clarity execrable, a pole vaulter's strength be splendid and his manners ill. On a trope analysis, the immediate object of evaluation is the trope, so that strictly speaking, different objects are being evaluated when we consider the flavor and the clarity of the wine, and thus the incompatible evaluations give rise to no problems at all. (Campbell, 1997 [1981), pp. 130 f. Cf. also Bacon, 1995, pp. 129-131.)

Whether value conflicts of the type Campbell envisages constitute any serious problem for an anti-tropist may well be doubted. After all, there is nothing strange in viewing an object as valuable insofar as it has one feature but disvaluable insofar as it has another. Such "pro tanto" evaluations of one and the same object can then be

---

[11] In a paper in which he argues that trope theory confronts formidable difficulties when it is called to explain what's going on when tropes themselves (and not just concrete objects) become subjects of predication, Fredrik Stjernberg concludes: "From an ontological perspective, tropes are simply *de trop*." (Stjernberg, 2003) He allows, however, that talk about tropes may still be quite useful in various theoretical contexts, in which the questions of fundamental ontology are not at issue. In particular, and in relation to our concerns, "[t]rope theory could be an interesting approach to problems in morally and aesthetically evaluating actions and objects." (ibid.) It is not clear, however, whether Stjernberg's worries about tropes as subjects of predication could not just as well apply to the predication of value with respect to tropes. On this issue, cf. Olson (2000).

[12] Note that even if the radical tropists were right in their suggestion that things and persons are just bundles of tropes, it would still be an open question whether the final values of such bundles are reducible to the final values of their various elements.

used as a basis for an overall evaluative assessment of the object in question. Still, Campbell's underlying suggestion that tropes are well equipped for the role of value bearers seems right. In fact, some tropes may well be considered to be valuable for their own sake. To take a simple case, *John's being pleased* may be something that has final value. Other examples might be John's being free or his being wise. More controversial examples would be the beautiful color of a certain object or a person's special form of humour.

So let us assume that tropes may well be finally valuable. On this assumption, can we go further and claim, as Williams suggests, that all evaluation is "focussed on abstracta"? In particular, is it true that the final value of a thing or a person, *a*, that is supposed to accrue to *a* in virtue of P, is "nothing but" the final value of the trope *being P* that occurs in *a*? This would mean that the final value of concrete objects is reducible to the final value of tropes.[13]

6. TROPICAL PROBLEMS

We are inclined to reject this reduction proposal. It does have merits, provided, of course, that we accept the existence of tropes to begin with. But it also has a number of serious weaknesses.

If the value of *a*, which accrues to *a* in virtue of its having P, were nothing but the value of a certain trope, viz., the instantiation of P in *a*, then one would expect that whenever a valuable trope occurs in an object, the valuation of a trope extends to the object, i.e., that the object itself is seen as *pro tanto* more valuable. However, this implication does not seem to hold in all cases. Thus, to take an example, suppose we value Mary's humour (where this valuation is of the appropriate sort, for its object's own sake). Do we thereby necessarily take any evaluative stand towards Mary herself? Do we thereby take her to be a more valuable person, for her own sake (and not just as a source of humour)? It doesn't seem so. Or to take another example, suppose we value, for its own sake, the colour of an object, say, its brilliant redness. Do we thereby necessarily value the object itself, for its own sake? Again, this implication need not hold. (In fact, that striking colour may make the object itself quite hideous!) Nor do we take John to be a more valuable person because of his pleasure just because we value his pleasure for its own sake. In general, it seems that when an object, *a*, is supposed to be finally valuable in virtue of its having P, then the trope, *a*'s having P, may also be assigned final value. But the opposite

[13] The conception of tropes as value bearers and the idea that the value of concrete objects might be reducible to the value of tropes is interestingly discussed in Olson (2000). While he got the idea from us, his treatment of the subject is largely independent and in several respects different from our own. Somewhat surprisingly, while Olson thinks that the reduction may succeed, he still believes that the success of the reduction would not obviate the need of assuming, at a fundamental level, final but non-intrinsic values. The reason is that, according to him, the final value of *tropes* is itself always supervenient on their relations to other tropes with which they are compresent. We have serious doubts on this score, but we prefer to leave the discussion of this issue to another occasion.

implication does not hold; *a*'s having P may sometimes be finally valuable without it being the case that this trope value has any relevance for the value of *a* itself.

But isn't this observation just an instance of the well-known Moorean insight that the value of a whole need not be an increasing function of the value of its parts? After all, for a radical tropist, a concrete object is a whole composed of a number of tropes. The value of such a whole need not increase with every increase in the value of its components.

That's true, of course. But the observation in question, however trite, still implies that the reduction proposal under consideration is unsatisfactory. If an instantiation of a property in an object may sometimes be valuable without thereby making the object valuable in that respect, then the reductionist's position appears to be untenable: the object's value, which accrues to it in virtue of its having a certain property, cannot be simply reduced to the value of the instantiation of that property in the object, i.e., to the value of the trope.

Another weakness of the reduction proposal under consideration has to do with the differences in pro-attitudes that are fitting with respect to different kinds of valuable objects. It seems that certain pro-attitudes that fit things or persons are not fitting or perhaps may even be impossible to hold with respect to tropes. Thus, for example, consider the attitude of respect. I can respect a person but it sounds odd, to say the least, to say that I respect a trope. I can respect Ann for her courage, but this is an attitude I hold towards Ann, and not towards her courage. I may well appreciate the latter or wish to show such courage myself, but respect is what I reserve for the person.[14] Honouring someone, because of his or her achievements, is another example of that same phenomenon. We appreciate the achievements, but we honour the achiever. Similarly, certain fitting responses towards valuable things, such as treasuring or protecting, do not seem appropriate as far as the corresponding tropes are concerned.

Conversely, some appropriate pro-attitudes towards tropes are not appropriate towards things or persons. For example, I can be exhilarated by a display of courage in Ann but it would be wrong to say that I am exhilarated by Ann, even though I may well come to value her more on account of her courage. Similarly, I can rejoice in her happiness, but I cannot rejoice in Ann herself.

These examples suggest that the value of a concrete object (of a person or a thing) cannot just consist in the value of the corresponding tropes. On our analysis, the value of any object is explicated in terms of the range of appropriate responses to the object in question. Consequently, the value of a concrete object and the value of a trope must be different from each other if the concrete object and the trope call for different responses.[15]

---

[14] Admittedly, "I respect your courage" is what one might say, sometimes. But we would suggest that an utterance like this is an expression of respect for a person. Ordinary language is flexible; we can easily convey our meaning by utterances that literally mean something else.

[15] Rønnow-Rasmussen (2002) makes a similar point with respect to instrumental value. Such value cannot accrue to tropes, simply because our attitudes towards tropes are very different from the ways in which we value an instrument.

But couldn't one argue that a pro-attitude such as, say, respect or admiration, which is not directed to a trope but to a person, can itself be analysed as a certain conglomerate of various pro-attitudes towards tropes? And couldn't the same apply to all other pro-responses that are called for by concrete objects (things or persons)? If such a reduction at the level of responses were viable, then this would again open up the possibility of the *value* reduction: the value of a concrete object could then be seen as nothing but the value of the tropes that are the objects of the responses that belong to the relevant conglomerate. Nothing we have said above excludes this possibility. But as long as such response reduction has not been carried out, nor even been seriously attempted, we remain sceptical as to whether this avenue is really viable.

We conclude, then, at least tentatively, that the reduction of the final value of things or persons to the final value of tropes, or states, seems to fail, after all. However, we want to emphasise, once again, that this conclusion should not be taken to mean that the former values cannot be derived from the latter. *If* tropes exist, then it may be the case that the final value of an object, *a*, which accrues to it in virtue of its having a property P, is in some sense derivable from the final value of that particular instantiation of P, i.e., from the final value of a certain trope or a certain constellation of tropes.[16] Alternatively, one might try to derive the value of a concrete object from the value of some states of affairs that concern that object. We prefer to leave these questions open, partly because we are unsure of the exact meaning of the claim that one value is derivable from another. Still, the question of derivability is a different issue from the one we have been addressing. Dependence of one value on another is not the same thing as reduction.

APPENDIX: STATES VS. FACTS AS BEARERS OF VALUE

Zimmerman (2001a) reports Noah Lemos' suggestion that a fact may be seen as the obtaining of a state of affairs, where the obtaining of a state of affairs is not the same as its mere existence (cf. Lemos 1994).[17] Insofar as we can refer to a state, such as, say, *John is happy*, that state exists. But it need not obtain; as things are, John might be a very unhappy person. If a state of affairs *S* obtains, the obtaining of *S* is a *fact*.[18] Facts are related to states more or less as tropes are to properties (universals). Just as a trope is an actual instantiation of a property, a fact is the obtaining of a state. In the same way as a property may exist without being instantiated, a state may exist without obtaining.

[16] According to many tropists, all tropes are simple entities. This would mean that, if P is a complex conjunction of the form $P = P_1 \& P_2 \&\ldots\& P_n$, then there is no such thing as one trope *a's being P*. Instead, we have a bundle of compresent tropes, *a*'s being P_1, *a*'s being P_2, etc.

[17] The distinction between states (Sachverhalte) and facts goes back at least to Wittgenstein's *Tractatus*.

[18] In Lemos (1994), facts were interpreted differently, as those states of affairs that obtain, rather than as their obtainings. On this point, judging from Zimmerman's report, Lemos now seems to have changed his view.

Zimmerman thinks that this characterization of facts as obtainings of states of affairs is on the right track, even though he suggests that the obtaining of a state of affairs may itself be analyzed as consisting in some individual (or individuals) exemplifying a certain property (relation) at a certain time. This latter suggestion is problematic,[19] but for our purposes the important point is that one way of looking at facts is to see them as the obtainings of states of affairs. Zimmerman then goes on to suggest, following Lemos (1994), that it is facts rather than states that are the bearers of value.[20] The mere existence of a state such as *John is happy*, does not make the world better; but the fact that John is happy does.

Actually, this argument does not seem to be decisive. That a state's existence does not make the world better only shows that the state lacks a *contributive* value as far as the value of the world is concerned, but it implies nothing about that state's value on its own, or for its own sake (cf. Danielsson 1999 and Olson 2000, appendix). To be sure, one might argue that the value of a state is *derived* from the value its obtaining would have (i.e., from the value of a fact), but a derivative character of values does not, as we have seen, automatically make them non-final. If final value is analysed in terms of fitting responses, then – it seems – it might be possible to ascribe final value to states, along with facts. To be sure, some responses towards finally valuable objects are clearly inappropriate with respect to states. We cannot, say, rejoice in a non-obtaining state (if we know it does not obtain), as we can in a fact. But other responses are not like this. Thus, a state may well be the intentional object of a wish, of a desire or of a preference. It is true that what we wish or desire is for the state to obtain, but it is the state itself, and not its obtaining,

---

[19] Zimmerman wants to extend this analysis even to existential facts such as the obtaining of the state that *there exists some x that has P*. If the x in question happens to be *a*, then this existential fact just consists in *a* exemplifying P. However, as Lemos has argued (in private communication, cf. Zimmerman 2001a, ch. 3), the obtaining of *there exists some x that has P* should not be identified with the obtaining of *a has P,* even when the former does obtain in virtue of the obtaining of the latter. For the two states involved are not logically equivalent. So the former fact, unlike the latter, does not consist in some individual exemplifying a certain property. Zimmerman resists this argument and opts for the view that the two facts in question may, after all, be identified with each other. At the same time, he argues for a *fine* individuation of facts, against "coarse" individuators. (For an analogous contrast concerning the individuation of events and actions, cf. Kim 1976 and Goldman 1970, on the "fine" side, and Anscombe 1969 and Davidson 1980, on the "coarse" side.). One might well wonder how this fine individuation of facts is supposed to square with his rejection of Lemos' argument. Furthermore, his reply to Lemos leaves some unanswered questions. Suppose there are many objects that have P. Would Zimmerman say that the obtaining of *there exists some x that has P* is in this case identical with the obtaining of a long conjunction *a has P and b has P and ...*? Or would he instead prefer to say that the state such as *there exists some x that has P* may have *many* distinct obtainings – as many, in fact, as there are objects that have *P*? In private communication, Zimmerman tells us that he would opt for the latter alternative. Be that as it may, he would still have problems with *universal* facts. What does the obtaining of a universal state *all x have P* consist in? Surely, this state does not have several distinct obtainings. Nor does its obtaining coincide with the obtaining of an (infinite) conjunction *a has P and b has P and ...*, even if the individuals that appear in this conjunction happen to be all the individuals that exist. For this conjunction by itself does not imply that no other individuals exist.

[20] This point also applies to his reduction proposal in Zimmerman (2001b) [* pp. 191-205 of this volume]: the value of *a*, which accrues to that object in virtue of its having P, is taken by him to be reducible to the value of the *fact* that *a* has P, and not to the value of the corresponding state of affairs.

towards which we have that attitude. Still, we may be wrong on this point: The value of states as opposed to facts is a complicated issue that need not be settled in this paper. In our discussion of various proposals how to reduce thing- or person-values to the values of states, we didn't bother about distinguishing facts from states. Our objections are applicable to both kinds of proposals alike. But it may turn out that Lemos and Zimmerman are right, after all, in their suggestion that it is facts rather than states that are proper value bearers. Still, as value bearers, facts are not alone: the final value of concrete objects is irreducible.

REFERENCES

Anderson E., (1993), *Value in Ethics and Economics*, Cambridge, Mass.: Harvard University Press.

Anscombe, G. E. M., (1969), *Intention*, Ithaca: Cornell University Press.

Baron Marcia W., (1997), "Kantian Ethics" in *Three Methods of Ethics*, Marcia W. Baron, Philip Pettit, and Michael Slote, Oxford: Blackwell.

Bacon, John (1997), "Tropes", in Zalta, E. N. (ed.), *Stanford Encyclopedia of Philosophy*, URL=<http://plato.stanford.edu/archives/fall2002/entries/tropes/>.

Bacon, John (1995), *Universals and Property Instances: The Alphabet of Being*, Aristotelian Society Series vol. 15, Oxford: Blackwell.

Brentano F., 1969 (1889). *The Origin of Our Knowledge of Right and Wrong*, London and Henley: Routledge & Kegan Paul.

Broad, C. D., (1930), *Five Types of Ethical Theory*, London: Keagan Paul, Trench, Trubner.

Brülde, Bengt (1998), *The Human Good*, Göteborg: Acta Universitatis Gothoburgensis.

Butchvarov, Panayot (1989), *Skepticism in Ethics*, Bloomington: Indiana University Press.

Campbell, Keith (1997), "The Metaphysic of Abstract Particulars", in Mellor, D.H. & Oliver, A., (eds.), *Properties*, Oxford University Press, pp. 125-139; reprinted from *Midwest Studies in Philosophy VI: The Foundations of Analytical Philosophy* (1981), pp. 477-488.

Campbell, Keith (1990), *Abstract Particulars*, Oxford: Basil Blackwell.

Danielsson, Sven (1999), "Om eget värde och annats", typescript, Uppsala University.

Davidson, Donald (1980), *Essays on Actions and Events*, Oxford: Clarendon Press.

Ewing, A. C., (1947), *The Definition of Good*, New York: MacMillan and Co.

Goldman, Alvin I., (1970), *A Theory of Human Action*, Princeton: Princeton University Press.

Kagan, Shelly, (1998), "Rethinking Intrinsic Value", *Journal of Ethics*, 2, pp. 277–97 [* pp. 97-114 of this volume].

Kim, Jaegwon (1976), "Events as Property Exemplifications", in Brand, M. and Walton, D. (eds.), *Action Theory*, Dordrecht: D. Reidel, pp. 159-77.

Lemos, Noah M., (1994), *Intrinsic Value: Concept and Warrant*, Cambridge: Cambridge University Press.

Olson, Jonas (2000), "The Bearers of Value", typescript, Lund University.

Rabinowicz, Wlodek & Rønnow-Rasmussen, Toni (1999), "A Distinction in Value: Intrinsic and For Its Own Sake", *Proceedings of the Aristotelian Society*, vol. 100, part 1, pp. 33-49 [* pp. 115-29 of this volume].

Rønnow-Rasmussen, Toni (2002), "Instrumental Values – Strong and Weak", *Ethical Theory and Moral Practice*, 5, pp. 23-43.

Ross, W. D., (1939), *Foundations of Ethics*, Oxford: Oxford University Press.

Stjernberg, Fredrik (2003), "An Argument against the Trope Theory", *Erkenntnis*, 59, pp. 37-46.
Swanton, Christine, (1995), "Profiles of the Virtues", *Pacific Philosophical Quarterly*, 27, pp. 335-44.
Williams, Donald C., (1966), *Principles of Empirical Realism*, Springfield: Charles C. Thomas.
Williams, Donald C., (1997), "The Elements of Being: I", in Mellor, D. H. & Oliver, A. (eds.), *Properties*, Oxford University Press, pp. 112-125, reprinted from *Review of Metaphysics*, 7 (1953), pp. 1-18.
Williams, Donald C., (1986) "Universals and Existents", *Australasian Journal of Philosophy*, 64, pp. 1-14.
Zimmerman, Michael, (2001a), *The Nature of Intrinsic Value*, Lanham: Rowman & Littlefield.
Zimmerman, Michael, (2001b), "Intrinsic Value and Individual Worth", in Dan Egonsson, Jonas Josefsson, Björn Petterson and Toni Rønnow-Rasmussen (eds.), *Exploring Practical Philosophy: From Action to Values*, Aldershot: Ashgate Publishers, pp. 123-38 [* pp. 191-205 of this volume].

PART IV

THE LOGIC OF INTRINSIC VALUE

CHAPTER 18

R. M. CHISHOLM

THE INTRINSIC VALUE IN DISJUNCTIVE STATES OF AFFAIRS

1. INTRODUCTION

I will propose and defend two general principles about the intrinsic value to be found in disjunctive states of affairs. These principles may be thought of as supplementing the logic of intrinsic value that has been developed by Ernest Sosa and me.[1]

2. SOME PHILOSOPHICAL PRESUPPOSITIONS

The intrinsic value of a state of affairs is the value that that state of affairs has necessarily. Hence, it is the value that the state of affairs has in every possible world in which it obtains. To assess the intrinsic value of a given state of affairs, we determine the amount of good and evil that the state of affairs guarantees to every possible world in which it obtains, and then we weigh its "best" against its "worst". More exactly, to the extent that we are able to assess the intrinsic value of a given state of affairs, (1) we ascertain its "best", i.e., the maximum amount of goodness the state of affairs guarantees to every possible world in which it obtains, (2) we ascertain its "worst", i.e., the maximum amount of evil it guarantees to every possible world in which it obtains, and then (3) we ascertain whether the first quantity is greater than, less than, or equal to the second.

This general point may be illustrated by reference to a hedonistic theory of value (though it is by no means tied to such a theory). In order to make certain logical points, I will suppose that hedonism is the correct theory of value: the only things that are intrinsically good are states of affairs implying that there is pleasure, and the only things that are intrinsically bad are states of affairs implying that there is displeasure. I will also suppose that pleasures and displeasures can be ordinally ranked, that the more pleasure the better and the less displeasure the better, and that equal amounts of pleasure and displeasure balance each other off.

---

[1] See Chisholm and Sosa [4] and the discussions of this system in Åqvist [1], Parsons [8], and Rescher [9].

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 229-239.
© 2005 *Springer. Printed in the Netherlands.*

In applying hedonistic criteria for assessing the intrinsic value of a state of affairs, we should consider three questions. (1) What is the maximum amount of pleasure ("the best") which is such that the state of affairs guarantees that amount of pleasure to every possible world in which it obtains? (2) What is the maximum amount of displeasure ("the worst") which is such that the state of affairs guarantees that amount of displeasure to every possible world in which it obtains? (3) Is the maximum amount of pleasure thus guaranteed greater than, less than, or equal to the maximum amount of displeasure thus guaranteed? If the maximum amount of pleasure is greater than that of displeasure, the state of affairs is good; if it is less, the state of affairs is bad; and if it is the same, the state of affairs is neither good nor bad.[2]

These procedures would seem to be obvious enough, but they lead to results that are incompatible with most theories of preferability.

Let us consider them in application to the following states of affairs:

(*p*) Jones experiencing 1 unit of pleasure,

(*q*) Smith experiencing 1 unit of displeasure,

(*r*) There being stones,

(*s*) Brown experiencing 2 units of pleasure,

(*t*) Black experiencing 2 units of displeasure.

Given our hedonistic assumptions, there is, of course, no problem in assessing the intrinsic value of these five states of affairs and in ranking them. Thus, *p* and *s* are good, *q* and *t* are bad, and *r* is neutral; and *s* is the best of the lot, and *t* is the worst. Moreover, there is no problem involved in evaluating the various possible conjunctions. Thus, *p&q* is neutral, *p&r* is good, *p&s* is good, and *p&t* is bad and has the same value as *q*. And the conjunction of all five is neutral.

But what of the negations of these states of affairs? According to many theories of value, the negation of a bad state of affairs is good, and the negation of a good state of affairs is bad. According to our present criteria, however, the negation of each of these states of affairs is neutral. For each negation may obtain in worlds in which there is no pleasure and no displeasure.

And what of the disjunctions of these states of affairs?

The disjunction $p \vee q$ (it being the case either that Jones experiences 1 unit of pleasure or that Smith experiences 1 unit of displeasure) is neutral. For it may obtain

[2] These questions have their analogues for nonhedonistic theories of intrinsic value. But these analogues, particularly that of the third question, are often difficult if not impossible to answer. Hence, given such theories, it may be impossible to assess the value of certain "mixed" states of affairs having both good and bad consequences.

in worlds in which there is no pleasure, and it may obtain in worlds in which there is no displeasure. Hence, the maximum amount of pleasure which is such that the state of affairs guarantees that amount of pleasure to every possible world in which it obtains is none at all, and analogously for displeasure. We may say, therefore, that $p \vee q$ falls between its disjuncts in value. And so, too, for $p \vee r$, $q \vee r$, $s \vee t$, $r \vee s$, $r \vee t$, $p \vee t$, and $q \vee s$.

Consider now $p \vee s$ – it being the case either that Jones experiences 1 unit of pleasure or that Brown experiences 2 units of pleasure. This is obviously intrinsically good, for it guarantees 1 unit of pleasure to every world in which it obtains, and it may obtain in worlds in which there is no displeasure. But it does not guarantee any more than 1 unit of pleasure to every world in which it obtains, for it may obtain in worlds in which there is only 1 unit of pleasure. Hence, the value of $p \vee s$ is the same as p. Thus, $p \vee s$ has the same value as its lesser-valued disjunct.

Analogously for $q \vee t$ – it being the case either that Smith experiences 1 unit of displeasure or that Black experiences 2 units of displeasure. This guarantees a minimum of 1 unit of displeasure to every possible world in which it obtains, and it may obtain in worlds in which there is no pleasure. Since it may obtain in worlds in which there is no more than 1 unit of displeasure, $q \vee t$ has the same value as t. Hence, $q \vee t$ has the same value as its higher-valued disjunct.

Castañeda has noted that there are three general methods for evaluating disjunctions: (a) the value of the disjunction is the same as that of the higher-valued disjunct, (b) the value of the disjunction is the same as the value of the disjuncts if these are the same in value, and otherwise it lies between the disjuncts in value, and (c) the value of the disjunction is the same as the value of its lower-valued disjunct.[3]

But if what we have just said is correct, then no one of these procedures is correct. For $p \vee q$ falls in value between its disjuncts and thus falsifies (a) and (c); of $p \vee r$ and $p \vee s$, each has the same value as its lesser-valued disjunct and thus falsifies (a) and (b); and of $q \vee r$ and $q \vee t$, each has the same value as its higher-valued disjunct and thus falsifies (b) and (c).

Such facts as these require that the principles of the logic of intrinsic preferability differ in fundamental respects from most other systems of the logic of preferability and of preference. And they require that, if we are to define "intrinsically good" and "intrinsically bad" in terms of "intrinsically preferable", we provide a set of definitions quite different from those set forth in the other systems.

3. A LOGIC OF INTRINSIC VALUE

I will now summarize informally the system of the logic of intrinsic value set forth by Sosa and me.

---

[3] See Castañeda [2] and cf. Freeman [6]. Castañeda adopts a recursive procedure involving rejection of all three methods, and Rescher, in [9], defends a version of (b). It should be emphasized that these authors have not restricted their principles to intrinsic value.

Taking the concept of intrinsic preferability as undefined and expressing "p is intrinsically preferable to *q*" as "*p*P*q*", we proposed these definitions:

(*D*1)	*p* has the same intrinsic value as *q*	$(p\mathrm{S}q) =_{Df} \sim(p\mathrm{P}q)\ \&\ \sim(q\mathrm{P}p)$.
(*D*2)	*p* is intrinsically indifferent	$(\mathrm{I}p) =_{Df} \sim(p\mathrm{P}{\sim}p)\ \&\ \sim({\sim}p\mathrm{P}p)$.
(*D*3)	*p* is intrinsically neutral	$(\mathrm{N}p) =_{Df} (\exists q)(\mathrm{I}q\ \&\ p\mathrm{S}q)$.
(*D*4)	*p* is intrinsically good	$(\mathrm{G}p) =_{Df} (\exists q)(\mathrm{I}q\ \&\ p\mathrm{P}q)$.
(*D*5)	*p* is intrinsically bad	$(\mathrm{B}p) =_{Df} (\exists q)(\mathrm{I}q\ \&\ q\mathrm{P}p)$.

I add this further definition to facilitate exposition:

(*D*6) *p* is at least as good intrinsically as *q* $(p\mathrm{A}q) =_{Df} \sim(q\mathrm{P}p)$.

The system includes axioms equivalent to the following:

(*A*1) $(p)(q)[p\mathrm{P}q \rightarrow \sim(q\mathrm{P}p)]$.

(*A*2) $(p)(q)(r)[(q\mathrm{A}p\ \&\ r\mathrm{A}q) \rightarrow r\mathrm{A}p]$.

(*A*3) $(p)(q)[(\mathrm{I}p\ \&\ \mathrm{I}q) \rightarrow p\mathrm{S}q]$.

(*A*4) $(p)[(\mathrm{G}p \vee \mathrm{B}{\sim}p) \rightarrow p\mathrm{P}{\sim}p]$.

We make use of rules corresponding to *modus ponens* and to a principle of substitution for tautological equivalents. The latter principle enables us to say that tautologically equivalent states of affairs are the same in value.

It is important to note the distinction between intrinsic indifference (I*p*) and intrinsic neutrality (N*p*). The class of the neutral is wider than that of the indifferent. A neutral state of affairs (e.g., there being no pleasure) may have a good negation. And a neutral state of affairs (e.g., there being no displeasure) may have a bad negation. But an indifferent state of affairs is a neutral state of affairs that has a neutral negation.

This system is adequate to what we said in our informal remarks about the relation between nonneutral states of affairs and their negations. But it contains no principles pertaining to disjunctions.

4. A PRINCIPLE ABOUT DISJUNCTIONS

The first of the two principles about disjunctions that I will propose may suggest itself if we reflect upon what was said informally above. We had noted that if we are hedonists, then in evaluating a given state of affairs (1) we will consider the

minimum amount of pleasure that state of affairs guarantees to every possible world in which it obtains, (2) we will consider the minimum amount of displeasure it guarantees to every possible world in which it obtains, and then (3) on the basis of the first two findings, we will estimate the value of the state of affairs.

If this is the correct procedure, then the value of the disjunction will never exceed the value of its higher-valued disjunct. For the minimum value the disjunction will guarantee to every possible world in which it obtains will be that of its higher-valued disjunct. And by similar reasoning we may also conclude that the value of the disjunction will never be lower than that of its lower-valued disjunct. And so we could say that the value of the disjunction *falls within the area of the values* of its disjuncts and give this expression a precise meaning.

We could say that a state of affairs *falls in value between the values* of two other states of affairs provided only that it is better than one and worse than the other. And we could say that a state of affairs *falls within the area of the values* of two states of affairs provided it is not the case that they are both better than it is or that they are both worse than it is. So whatever falls in value between the values of two states of affairs also falls within the area of the values of those states of affairs. But the converse will not be true, for a state of affairs may fall within the area of the values of two states of affairs and be the same in value as one of them or as both of them.

Our first axiom pertaining to disjunctive states of affairs, then, will tell us that every disjunction falls within the area of the values of its disjuncts. In other words:

$$(A5) \qquad \sim [p\mathrm{P}(p \vee q)\ \&\ q\mathrm{P}(p \vee q)]\ \&\ \sim[(p \vee q)\mathrm{P}p\ \&\ (p \vee q)\mathrm{P}q].$$

And this means that every disjunction is such that it is at least as good intrinsically as one of its disjuncts, and one of its disjuncts is at least as good intrinsically as it is. For an obvious consequence of (*A*5) and the definition of "at least as good as" ("$p\mathrm{A}q$") is this:

$$(T1) \qquad [p\mathrm{A}(p \vee q) \vee q\mathrm{A}(p \vee q)]\ \&\ [(p \vee q)\mathrm{A}p \vee (p \vee q)\mathrm{A}q].$$

We could say, in still other words, that no disjunction is such that both disjuncts are better than it is or both disjuncts are worse than it is. Or again: no disjunction is better than its better disjunct or worse than its worse disjunct.

We can now derive two further principles, versions of which are sometimes taken as axioms in preference logic (cf. Hansson [5]: 20). The first is:

$$(T2) \qquad r\mathrm{P}(p \vee q) \rightarrow (r\mathrm{P}p \vee r\mathrm{P}q).$$

This principle tells us that whatever is better than a given disjunction is better than one of the disjuncts of that disjunction. And that it follows from the foregoing may be seen this way: suppose (*T*2) were not the case; i.e., suppose that r is better than a

certain disjunction and that each disjunct of that disjunction is at least as good as r. Then each disjunct would be better than the disjunction, and this contradicts ($T1$).

The second principle that is sometimes taken as an axiom in preference logic is this:

$$(T3) \qquad (p \vee q)\mathrm{P}r \rightarrow (p\mathrm{P}r \vee q\mathrm{P}r).$$

This tells us that whatever is worse than a given disjunction is also worse than one of the disjuncts of that disjunction. Suppose this were not the case; i.e., suppose that r is worse than the disjunction and r is at least as good as each of the disjuncts of that disjunction. Then the disjunction would be better than each of its disjuncts, and this contradicts ($T1$).

We note, for future reference, two further consequences of ($A5$):

$$(T4) \qquad (p\mathrm{P}q) \rightarrow [p\mathrm{A}(p \vee q) \ \& \ (p \vee q)\mathrm{A}q].$$

$$(T5) \qquad (p\mathrm{S}q) \rightarrow [(p \vee q)\mathrm{S}q].$$

In each case, the derivation is obvious. These latter principles have also been taken as basic in preference logic.[4]

5. A SECOND PRINCIPLE ABOUT DISJUNCTIONS

The disjunctions considered up to now are relatively simple, even within the framework of our (fictitious) hedonistic assumptions, for they do not contain what Moore called "mixed goods" and "mixed evils" (cf. [7], Ch. 6). We could say, in the spirit of Moore: a *mixed good* is a good state of affairs that entails a bad state of affairs, a *mixed evil* is a bad state of affairs that entails a good state of affairs, a *pure good* is a good state of affairs that is not mixed, a *pure evil* is a bad state of affairs that is not mixed, a *mixed neutral* is a neutral state of affairs that either entails a bad state of affairs or entails a good state of affairs, and a *pure neutral* is a neutral state of affairs that is not mixed.

Our first principle pertaining to disjunctions enables us to deal in a straightforward way with disjunctions having as disjuncts only pure goods, or pure evils, or pure neutrals. But it does not enable us to deal with disjunctions having disjuncts that are mixed. Yet the intuitive considerations we have already set forth make clear what we should say about such disjunctions.

Let us consider a single example and view it once more within the context of our assumed hedonistic theory of value. The example will be a disjunction of a mixed good and a mixed evil, and one in which the amount of good guaranteed by one

---

[4] Cf. Castañeda [2]: 264: "The value of a utilitarian disjunction cannot exceed the value of its highest-valued disjunct" and "The value of a utilitarian disjunction cannot be exceeded by the value of its lowest-valued disjunct."

disjunct is not the same as that guaranteed by the other and in which the amount of evil guaranteed by one disjunct is not the same as that guaranteed by the other:

> (Someone experiencing 5 units of pleasure and someone experiencing 3 units of displeasure) or (someone experiencing 2 units of pleasure and someone experiencing 1 unit of displeasure).

This disjunction, if it obtains, guarantees the universe both good and evil. It will guarantee as much good as would be guaranteed by the disjunct that guarantees the lesser amount of good; for the amount of good guaranteed by *that* disjunct will be assured whichever disjunct obtains. Hence, the disjunction guarantees 2 units of pleasure. And it will guarantee as much evil as would be guaranteed by the disjunct that guarantees the lesser amount of evil; for the amount of evil guaranteed by *that* disjunct will be assured whichever disjunct obtains. Hence, the disjunction, guarantees 1 unit of displeasure.

Note that we are *not* saying that any such disjunction guarantees as much good as is guaranteed by its worse disjunct, or that it guarantees as much evil as is guaranteed by its better disjunct. Thus, the disjunction of our example does not guarantee as much evil as is guaranteed by its better disjunct (for the disjunction guarantees 1 unit of displeasure and its better disjunct guarantees 3). But this particular disjunction does guarantee as much evil (1 unit of displeasure) as is guaranteed by its worse disjunct.

If we switch "pleasure" and "displeasure" in our statement of the example, we formulate a disjunction which does not guarantee as much good as is guaranteed by its worse disjunct, but which does guarantee as much good as is guaranteed by its better disjunct.

Generalizing on what we have said about the particular example, we may first give an approximate statement of our second disjunctive principle: any disjunction of goods and evils has the same value as does any state of affairs which (a) guarantees as much good as is guaranteed by the disjunct that guarantees the lesser amount of good and (b) guarantees as much evil as is guaranteed by the disjunct that guarantees the lesser amount of evil.

The thought behind our principle may be suggested by this somewhat oversimplified statement of it: "To find the relation between the value of a disjunction, $p \vee q$, and the values of its disjuncts, (1) consider the worst entailed by p and the worst entailed by q and then take whichever one, r, is at least as good as the other, (2) consider the best entailed by p and the best entailed by q and then take whichever one, s, is not better than the other; then $p \vee q$ has the same value as any t which is such that the worst in t has the same value as r and the best in t has the same value as s."[5]

In our more precise formulation of the principle, we will make use of the abbreviations "one of the bests in *p*" and "one of the worsts in p":

[5] I formulated this principle in [3]: 268 [* pp. 178-79 of this volume].

(*D*7) *q* is one of the bests in $p =_{Df}$ (i) *p* logically implies *q*, and (ii) for every *r*, if *p* logically implies *r*, then *q* is at least as good as *r*.

(*D*8) *q* is one of the worsts in $p =_{Df}$ (i) *p* logically implies *q*, and (ii) for every *r*, if *p* logically implies *r*, then *r* is at least as good as *q*.

The bests in *p* need not be good, and the worsts in *p* need not be bad. (But if, as seems plausible, we assume that every logically necessary state of affairs is neutral, then if the bests in *p* are not good, they will be neutral, and if the worsts in *p* aren't bad, they will be neutral.) Obviously, the bests in *p* are at least as good as *p*, and *p* doesn't entail anything that is better than they are. And *p* is at least as good as any of the worsts in *p*, and *p* doesn't entail anything worse than they are.

We may now formulate more precisely our second axiom pertaining to the intrinsic value in disjunctive states of affairs:

(*A*6) For every *p*, *q*, *r*, *s*, and *t*, if (a) *r* is a worst in *p* or a worst in *q*, and no worst in *p* or worst in *q* is better than *r*, (b) *s* is a best in *p* or a best in *q*, and no best in *p* or best in *q* is worse than *s*, and (c) the worsts in *t* have the same value as *r*, and the bests in *t* have the same value as *s*, then $p \vee q$ has the same value as *t*.

Let us now apply (*A*6) to the disjunction $p \vee q$ just discussed. We have:

(*p*) Someone experiencing 5 units of pleasure and someone experiencing 3 units of displeasure.

(*q*) Someone experiencing 2 units of pleasure and someone experiencing 1 unit of displeasure.

(*r*) Someone experiencing 1 unit of displeasure.

(*s*) Someone experiencing 2 units of pleasure.

(*t*) Someone experiencing 1 unit of displeasure and someone experiencing 2 units of pleasure.

Contemplating these five states of affairs and still presupposing our simplified hedonism, we can see that: *r* is the better of the worst two states of affairs; *s* is the worse of the best two states of affairs; and the worsts in *t* have the same value as *r*, and the bests in *t* have the same value as *s*.[6] Application of (*A*6), then, to our disjunctive state of affairs $p \vee q$, viz.,

[6] In this particular example, *t* is the conjunction of *r* and *s*. But (*A*6) does not require that *t* be the conjunction of *r* and *s*, and, as will be noted below (in replying to the second of the objections to (*A*6)), there are situations in which (*A*6) precludes *t* being the conjunction of *r* and *s*.

> (Someone experiencing 5 units of pleasure and someone experiencing 3 units of displeasure) or (someone experiencing 2 units of pleasure and someone experiencing 1 unit of displeasure),

tells us that that state is the same in value as the following state of affairs (*t*):

> Someone experiencing 1 unit of displeasure and someone experiencing 2 units of pleasure,

and hence that it is a "mixed good".

6. AN OBJECTION CONSIDERED

Consider now the following objection to what I have said: "(1) You are trying to formulate a way of computing the value of a disjunction from the values of its disjuncts. But (2) you can't find out the value of a state of affairs until you know the value of each of the states of affairs it entails. And (3) since each disjunct of a disjunction entails the disjunction itself, you can't find out the value of either disjunct until you've first found out the value of the disjunction itself. Therefore, (4) your procedure is circular."

To reply to this objection, we have only to note that the first premise is false. We are *not* trying to formulate "a way of computing the value of a disjunction from the values of its disjuncts." We are presupposing that we have a theory of value which, in accordance with the philosophical presuppositions noted at the outset, enables us to assess the intrinsic value of any given state of affairs. What we are trying to do is to formulate certain general principles relating the value to be found in any disjunctive state of affairs to the value to be found in each of its disjuncts.

But let us note in passing that the second premise of the objection is also false. To see that this is so, consider an analogous objection to the truth-table method of calculating the truth-value of a disjunction: "(1) You are trying to formulate a way of calculating the truth-value of a disjunction from the truth-values of its disjuncts. But (2) you can't find out the truth-value of a proposition until you know the truth-value of each of the propositions it entails. And (3) since each disjunct of a disjunction entails the disjunction, you can't find out the truth-value of either disjunct until you've first found out the truth-value of the disjunction itself. Therefore, (4) your method is circular."

The reply to the second argument is to deny the epistemological premise: "You can't find out the truth-value of a proposition until you know the truth-value of each of the propositions it entails." And so, too, for the first argument.

7. A PUZZLE SOLVED

I note finally the use of (*A*6) in solving a problem posed by Åqvist.

Åqvist considers the two following principles which we have derived from (*A*4):

($T2$) $r\mathrm{P}(p \vee q) \rightarrow (r\mathrm{P}p \vee r\mathrm{P}q)$.

($T3$) $(p \vee q)\mathrm{P}r \rightarrow (p\mathrm{P}r \vee q\mathrm{P}r)$.

He suggests that these principles, when taken together with the principle of substitution of tautological equivalents, which I have presupposed, lead to unacceptable consequences. Thus, we may readily prove

($T6$) $(p \vee q)\mathrm{P}r \rightarrow \{[p \vee (q \& s)]\mathrm{P}r \vee [q \vee (p \& s)]\mathrm{P}r\}$.

But ($T6$), Åqvist suggests, is counterintuitive, and he is inclined to think that the best way to avoid it is to reject the principle of substitution of tautological equivalents. Rejection of this principle, however, would seem too drastic a course – if we are actually concerned with the intrinsic value of states of affairs and not merely with the ways in which people may happen to evaluate states of affairs.

In objecting to ($T6$), Åqvist writes: "Let p be 'I am enjoying a drink of whiskey,' let q be 'I am enjoying a drink of brandy,' let r be 'I have a drink of milk,' and let s be 'I am suffering from a drink of cyanide'" ([1]: 264). But let us formulate his point using a different example in order not to lose sight of the fact that we are here concerned with intrinsic preferability and not with some other preference relation. Presupposing hedonism once again, let us consider these states of affairs:

(p) Jones having 5 units of pleasure.

(q) Smith having 2 units of pleasure.

(r) There being stones.

(s) Robinson having 100 units of displeasure.

According to the value scheme we are presupposing, *Jones having 5 units of pleasure or Smith having 2 units of pleasure* is intrinsically preferable to *there being stones.* In other words,

(i) $(p \vee q)\mathrm{P}r$.

Therefore, given ($T6$) above we can deduce

(ii) $[p \vee (q \& s)]\mathrm{P}r \vee [q \vee (p \& s)]\mathrm{P}r$.

But now, it is contended, (ii) is false. "Surely," one may say, "that state of affairs which is either Jones having 5 units of pleasure or it being the case both that Smith has 2 units of pleasure and that Robinson has 100 units of displeasure is not a state of affairs which, given the value system we are presupposing, is preferable to there being stones. Nor is that state of affairs which is either Smith having 2 units of

pleasure or it being the chase that Jones has 5 units of pleasure and that Robinson has 100 units of displeasure one that is preferable to there being stones."

But let us apply (*A*6) to the $p\vee(q\&s)$ of the example. The result is that $p\vee(q\&s)$ is good. The first disjunct, *p*, is *Jones having 5 units of pleasure*, and the second disjunct, *q&s*, is *Smith having 2 units of pleasure and Robinson having 100 units of displeasure.* One of the bests in *p*, then, will be *someone having 5 units of pleasure*, and one of the bests in *q&s* will be *someone having 2 units of pleasure.* The worst in *p* will be neutral, and a worst in *q&s* will be *someone having 100 units of displeasure.* Of the worsts in *p* and the worsts in *q&s*, we take that one, *r*, which is not worse than the other; hence, *r* will be neutral. Of the bests in *p* and the bests in *q&s*, we take that one, *w*, which is not better than the other; *w*, then, will be *someone having 2 units of pleasure.* Our whole disjunction now will have the same value as any state of affairs *t* which is such that the worst in *t* has the same value as the neutral *r* and the best in *t*. But *someone having 2 units of pleasure* is itself such a *t*. Since *t* is good, application of our general principle to the disjunctive state of affairs, $p\vee(q\&s)$, shows that it, too, is good. A similar result, of course, will hold for $q\vee(p\&s)$.

I conclude, then, that the proposed example is not counterintuitive.[7]

REFERENCES

[1] Lennart Åqvist, "Chisholm-Sosa Logics of Intrinsic Betterness and Value," *Noûs* 2 (1968): 253-70.

[2] Hector-Neri Castañeda, "Ought, Value, and Utilitarianism," *American Philosophical Quarterly* 6 (1969): 257-75.

[3] Roderick M. Chisholm, "Objectives and Intrinsic Value," in *Jenseits von Sein und Nichtsein*, ed. by Rudolf Haller (Graz: Akademische Druck- u. Verlagsanstalt, 1972): 261-69 [* pp. 171-79 of this volume].

[4] Roderick M. Chisholm and Ernest Sosa, "On the Logic of 'Intrinsically Better'," *American Philosophical Quarterly* 3 (1966): 244-49.

[5] Bengt Hansson, *Topics in the Theory of Preference Relations* (mimeographed by the Department of Philosophy, University of Lund, Sweden).

[6] James B. Freeman, "Fairness and the Value of Disjunctive Actions," *Philosophical Studies* 24 (1973): 105-11.

[7] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903).

[8] Charles Parsons, "Axiomatization of Åqvist's CS Logics," *Theoria* 36 (1970): 43-64.

[9] Nicholas Rescher, "Semantic Foundations for the Logic of Preference," in *The Logic of Decision and Action*, ed. by N. Rescher (Pittsburgh. University of Pittsburgh Press, 1967): 37-62.

[7] I am indebted to Robert Keim, Ernest Sosa, and the referee for many helpful suggestions. This paper was prepared in part under NSF Grant GS-2953.

CHAPTER 19

P. L. QUINN

IMPROVED FOUNDATIONS FOR A LOGIC OF INTRINSIC VALUE

According to Roderick M. Chisholm, the intrinsic value of a state of affairs is the value that state of affairs has necessarily. Hence, it is the value the state of affairs has in every possible world where it obtains. What logical principles may we assume to govern the conception of intrinsic value so understood?

1. A LOGIC OF INTRINSIC VALUE

In his most recent discussion of this topic Chisholm proposes a system of logic for intrinsic value. We begin with the primitive locution '*p* is intrinsically preferable to *q*', which is abbreviated '*pPq*'. In terms of this primitive we define the following locutions:

(D1) *p* has the same intrinsic value as *q*
pSq = Df $\sim(pPq)$ & $\sim(qPp)$

(D2) *p* is intrinsically indifferent
Ip = Df $\sim(pP{\sim}p)$ & $\sim({\sim}pPp)$

(D3) *p* is intrinsically neutral
Np = Df $(\exists q)(Iq \;\&\; pSq)$

(D4) *p* is intrinsically good
Gp = Df $(\exists q)(Iq \;\&\; pPq)$

(D5) *p* is intrinsically bad
Bp = Df $(\exists q)(Iq \;\&\; qPp)$

(D6) *p* is at least as good intrinsically as *q*
pAq = Df $\sim(qPp)$.

To these definitions we add the following axioms:

(Al) $(\forall p)(\forall q)(pPq \rightarrow \sim(qPp))$

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 241-248.
© 2005 *Springer. Printed in the Netherlands.*

(A2) $(\forall p)(\forall q)(\forall r)((qAp \,\&\, rAq) \rightarrow rAp)$

(A3) $(\forall p)(\forall q)((Ip \,\&\, Iq) \rightarrow pSq)$

(A4) $(\forall p)((Gp \vee B{\sim}p) \rightarrow pP{\sim}p)$

(A5) $(\forall p)(\forall q)({\sim}(pP(p\vee q) \,\&\, qP(p\vee q)) \,\&\, {\sim}((p\vee q)Pp \,\&\, (p\vee q)Pq))$.

Chisholm supplements these axioms with a semi-formalized principle about disjunctions having disjuncts which are mixed goods, mixed bads or mixed neutrals. Since nothing I will say in this paper bears on that principle, it will not be stated here.[1] This system of logic is appropriately thought of as an extension of a system proposed earlier by Chisholm and Ernest Sosa.[2]

A grammatical point is worth noting. Because Chisholm interprets the ordinary propositional variables as ranging over states of affairs, they are noun-surrogates rather than sentence-surrogates. Consequently, the connectives do double duty. In some occurrences they operate upon nouns to produce other nouns; in other occurrences they operate upon sentences to produce other sentences. This equivocation is harmless because it can always be made clear in a given context how a given occurrence of a connective is to be understood.

2. AN OBJECTION FORMULATED

It seems that there are states of affairs which are incomparable (incommensurable) with respect to intrinsic value. To say of states of affairs p and q that they are incomparable with respect to intrinsic value is to say that ${\sim}(pPq)$ and ${\sim}(qPp)$ and ${\sim}(pSq)$. To illustrate the general point using a hedonistic theory of value, let us consider the following three states of affairs:

(p_1) Smith enjoying the taste of apples

(q_1) Smith enjoying the sound of Beethoven's Ninth Symphony

(r_1) Smith enjoying the taste of pears.[3]

It may be that p_1 is intrinsically preferable to r_1 but that neither p_1 nor r_1 is intrinsically comparable to q_1. If this were so, the following relations would hold among these three states of affairs: p_1Pr_1; ${\sim}(p_1Pq_1)$; ${\sim}(q_1Pp_1)$; ${\sim}(p_1Sq_1)$; ${\sim}(q_1Pr_1)$; ${\sim}(r_1Pq_1)$; and ${\sim}(r_1Sq_1)$.

---

[1] The definitions and axioms I have stated are to be found on pp. 298-300 of [1] [* pp. 232-33 of this volume]. The axiom I have omitted is formulated on p. 303 of [1] [* p. 236 of this volume].

[2] The logic of intrinsic value formulated by Chisholm and Sosa in [2] contains (D1)-(D5) but not (D6). It contains axioms equivalent to (Al)-(A4) but does not contain (A5).

[3] This example is used by Hansson to make a related point on p. 431 of [3]. However, because Hansson is discussing the logic of preference, the import of the example for the logic of intrinsic value is not immediately evident from what he says there.

If we suppose that these relations do hold among p_1, q_1 and r_1, then Chisholm's logic is demonstrably inadequate in at least two ways. First, an immediate consequence of (D1) is this:

$$(\forall p)(\forall q)((\sim(pPq)\ \&\ \sim(qPp)) \rightarrow pSq).$$

But both p_1 and q_1 and q_1 and r_1 are, given our supposition, counterexamples to this principle. Second, given (D6), (A2) can be rewritten as follows:

$$(\forall p)(\forall q)(\forall r)((\sim(pPq)\ \&\ \sim(qPr)) \rightarrow \sim(pPr)).$$

And p_1, q_1 and r_1 are, on our supposition, a counterexample to this principle.

In assessing the gravity of this objection to Chisholm's logic of intrinsic value, we must ultimately take a stand on whether there are intrinsically incomparable states of affairs. Thus, we must ask ourselves: are there really states of affairs which are incomparable with respect to their intrinsic value? And, in trying to answer this question, we must also come to grips with a question about philosophical methodology: how is it to be determined whether there are states of affairs incomparable with respect to their intrinsic value? In the next section I will propose an answer to the methodological question, and then in the section after that I will employ my methodological proposal to generate what I take to be a satisfactory answer to the original question.

3. METHODOLOGICAL CONSIDERATIONS

I assume that it is genuinely problematical, philosophically speaking, whether or not there are states of affairs incomparable with respect to their intrinsic value. I take this assumption to mean that we are not entitled, on the basis of our intuitive and pre-theoretical judgments alone, to assume either that there are such states of affairs or that there are not such states of affairs. Hence, on that basis alone, we are not warranted in assuming that the supposition of the preceding section generates a genuine counterexample to Chisholm's logical principles and we are not warranted in assuming that it does not. In such a situation, we ought to allow our theory of intrinsic value to settle for us the question of whether there are intrinsically incomparable states of affairs. How else, after all, could the question be settled if it is of this sort?

But our theory ought not to settle the question by begging it. One way in which a theorist would beg the question at issue would be by defining things in such a way that there could not be intrinsically incomparable states of affairs. Since Chisholm's (D1) has just this effect, his theory does beg this important question and is, therefore, philosophically inadequate.[4] Another way in which a theorist would beg

[4] The corresponding definition in the system set forth by Chisholm and Sosa in [2] suffers from the same infirmity.

the question at issue would be by merely postulating that any two states of affairs are comparable with respect to their intrinsic value. This procedure would have, as Russell has noted, all the advantages of theft over honest toil.[5] Thus, if a theory is to settle the question which concerns us without begging it, that theory must contain a proof that there are states of affairs incomparable with respect to their intrinsic value, or a proof that there are no such states of affairs, from definitions and axioms such that each one of them leaves the question at issue open and is warranted by considerations other than those directly relevant to settling the question at issue.

We should take care not to confuse this methodological point with an obvious but unhelpful terminological point. Someone working on the logic of preference, for example, is free to stipulate that he understands the term ‘preference relation’ in such a way that a preference relation is definitionally transitive and strongly connected.[6] But, then, it is an open question whether or not the relation which holds between two objects of choice just in case a given individual prefers one to the other is a preference relation in the specified sense. If, as a matter of fact, that individual’s preferences are not transitive, then that relation is not a preference relation as defined by our theorist. Similarly, Chisholm could stipulate that he takes the term ‘intrinsically preferable’ to be defined in such a way that (D1)-(D6) and (Al)-(A5) cannot be false. But, if he were to do so, then it would be an open question whether or not the relation which holds between two states of affairs just in case the intrinsic value of one is greater than the intrinsic value of the other is the relation of intrinsic preferability so defined. And, if, as it happens, there are states of affairs incomparable with respect to intrinsic value, then the relation which holds among them is not the relation so defined.

4. ANOTHER LOGIC OF INTRINSIC VALUE

Let us begin anew with ‘*p* is at least as good intrinsically as *q*’, abbreviated as ‘pAq’, as our primitive locution. We may then lay down the following definitions:

(d1) *p* has the same intrinsic value as *q*

$$pSq =_{\mathrm{Df}} pAq \;\&\; qAp$$

(d2) *p* is intrinsically preferable to *q*

$$pPq =_{\mathrm{Df}} pAq \;\&\; {\sim}(qAp)$$

(d3) *p* is intrinsically incomparable to *q*

$$pUq =_{\mathrm{Df}} {\sim}(pAq) \;\&\; {\sim}(qAp).$$

These definitions are both exclusive and exhaustive in the sense that for any *p* and *q* exactly one of the following four relations holds between them: pSq, pPq, qPp or

[5] Russell’s epigram is found on p. 71 of [5].

[6] This is, for instance, the usage adopted by Hansson in [3] on p. 426 and in [4] on p. 443, as well as by some other theorists.

pUq. To these definitions we may add others, which capture the remaining concepts in Chisholm's logic of intrinsic value, as follows:

(d4) *p* is intrinsically indifferent
$Ip =\text{Df}\ pA{\sim}p \ \&\ {\sim}pAp$

(d5) *p* is intrinsically neutral
$Np = \text{Df}\ (\exists q)(Iq \ \&\ pSq)$

(d6) *p* is intrinsically good
$(Gp) = \text{Df}\ (\exists q)(Iq \ \&\ pPq)$

(d7) *p* is intrinsically bad
$Bp = \text{Df}\ (\exists q)(Iq \ \&\ qPp)$.

It is clear that (dl)-(d7) do not beg the question about whether or not there are states of affairs incomparable with respect to their intrinsic value. These definitions do not, together with propositional logic, imply that there are such states of affairs, nor do they imply that there are not such states of affairs. We might express this point by saying that, logically speaking, our definitions permit but do not require that there are intrinsically incomparable states of affairs but they also permit but do not require that there are no such states of affairs. Hence, (d1)-(d7) are in accord with the methodological doctrine set forth in the preceding section.

From (d2), together with propositional logic, we may easily deduce the following theorem:

(t1) $(\forall p)(\forall q)((pPq) \rightarrow {\sim}(qPp))$.

Therefore, our system of logic can do without an axiom analogous to (Al) in Chisholm's system. The analogues of Chisholm's (A2)-(A4) may be accepted for our purposes without argument:

(a1) $(\forall p)(\forall q)(\forall r)((pAq) \ \&\ qAr) \rightarrow pAr)$

(a2) $(\forall p)(\forall q)((Ip \ \&\ Iq) \rightarrow pSq)$

(a3) $(\forall p)((Gp \vee B{\sim}p) \rightarrow pP{\sim}p)$.

It seems evident that all these axioms express reasonable assumptions concerning intrinsic value. For instance, although (al) may not express a truth about empirical preferences, this provides no reason at all for thinking that it is not a truth about intrinsic preferability. And none of these axioms violates the methodological doctrine we have adopted, for none of them implies that there are intrinsically incomparable states of affairs or that there are no such states of affairs.

But what are we to say, in this system, about disjunctive states of affairs? The formal analogue of Chisholm's axiom (A5) would be satisfied even in the case of

some p and q such that $p \vee q$ is intrinsically incomparable with both its disjuncts. Surely, then, we are entitled to assume something stronger than this, namely, that any disjunction is comparable in intrinsic value with at least one of its disjuncts. Moreover, it also seems safe to assume, along with Chisholm, that every disjunction is such that at least one of its disjuncts is at least as good intrinsically as it is and it is at least as good intrinsically as at least one of its disjuncts is. This much can be guaranteed if we postulate the following principle:

$$\text{(a4)} \qquad (\forall p)(\forall q)(pA(p \vee q) \vee qA(p \vee q)) \,\&\, ((p \vee q)Ap \vee (p \vee q)Aq)).$$

The axiom (a4) by itself would hold even in the case of some p and q such that $p \vee q$ has the same intrinsic value as one of its disjuncts and is intrinsically incomparable to the other. Therefore, in adopting it, we do not deviate from our methodological doctrine. But, because disjunction is a symmetric relation, a disjunction should be intrinsically comparable to both its disjuncts or to neither. This can be guaranteed if we make the following assumption:

$$\text{(a5)} \qquad (\forall p)(\forall q)(pA(p \vee q) \vee (p \vee q)Ap) \leftrightarrow (qA(p \vee q) \vee (p \vee q)Aq)).$$

The axiom (a5) by itself would hold even in the case of some p and q such that $p \vee q$ is intrinsically incomparable to both of its disjuncts. And so we do not deviate from our methodological precepts in accepting it either. However, (a4) and (a5) together suffice to guarantee that any disjunction is comparable in intrinsic value to both its disjuncts.

From our axioms (al), (a4) and (a5) we are also able to infer that every pair of states of affairs is intrinsically comparable. A sketch of the proof begins with the observation that (a5) is equivalent to the following formula:

$$(\forall p)(\forall q)(((pA(p \vee q) \vee (p \vee q)Ap) \,\&\, (qA(p \vee q) \vee (p \vee q)Aq)) \vee \\ (\sim(pA(p \vee q)) \,\&\, \sim((p \vee q)Ap) \,\&\, \sim(qA(p \vee q)) \,\&\sim((p \vee q)Aq))).$$

The second disjunct of this formula is incompatible with (a4). Hence, the conjunction of (a4) and (a5) implies the formula:

$$(\forall p)(\forall q)((pA(p \vee q) \vee (p \vee q)Ap) \,\&\, (qA(p \vee q) \vee (p \vee q)Aq) \,\&\, \\ (pA(p \vee q) \vee qA(p \vee q)) \,\&\, ((p \vee q)Ap \vee (p \vee q)Aq)).$$

Using the distributive laws, we can expand this formula into a disjunction having sixteen disjuncts, each of which is a conjunction having four conjuncts. We discover by inspection of these sixteen disjuncts that each, together with (a1), implies either that pAq or that qAp. Hence, each implies that $pAq \vee qAp$. We thus obtain the following theorem:

(t2) $(\forall p)(\forall q)(pAq \vee qAp)$.

In the light of (d3), what (t2) tells us is that no two states of affairs are intrinsically incomparable. In other words, the relation expressed by '*pAq*' is strongly connected. Thus we have succeeded in settling our original question without exploiting the advantages theft has over honest toil and without violating the methodological precepts we adopted in the preceding section. Taken together, (al) and (t2) suffice to show that the relation expressed by '*pAq*' is a weak ordering or total preordering of all states of affairs. That relation is therefore a 'preference relation' in the sense of the term adopted by Hansson in [3] and [4]. The only surprising feature of this result is the fact that the conclusion that the relation in question is strongly connected follows from such intuitively obvious assumptions about disjunctive states of affairs as (a4) and (a5) when they are coupled with (al).

5. FURTHER RESULTS

It can also be shown that Chisholm's logic of intrinsic value is properly contained in the system we have been engaged in constructing. In order to prove this fact, we need to call attention to four additional consequences of our logic. The first is this:

(t3) $(\forall p)(\forall q)(\sim(pP(p\vee q) \,\&\, qP(p\vee q)) \,\&\, \sim((p\vee q)Pp \,\&\, (p\vee q)Pq))$.

Suppose (t3) were false for some p and q. For those particular states of affairs, the following formula would hold:

$$(pP(p\vee q) \,\&\, qP(p\vee q)) \vee ((p\vee q)Pp \,\&\, (p\vee q)Pq).$$

In the light of (d2), the first disjunct of this formula implies

$$\sim((p\vee q)Ap) \,\&\, \sim((p\vee q)Aq)$$

which contradicts the second conjunct of (a4). And its second disjunct implies

$$\sim(pA(p\vee q)) \,\&\, \sim(qA(p\vee q))$$

which contradicts the first conjunct of (a4).

Another theorem in our logical system is this:

(t4) $(\forall p)(\forall q)(pAq \leftrightarrow \sim(qPp))$.

The proof of the implication from left to right uses only (d2) plus propositional logic; however, the proof of the implication from right to left also makes use of (t2). From (t4) we can infer the following:

(t5) $(\forall p)(\forall q)((pAq \,\&\, qAp) \leftrightarrow (\sim(qPp) \,\&\, \sim(pPq)))$.

And an immediate consequence of (t5) is this:

(t6) $(\forall p)((pA{\sim}p \,\&\, {\sim}pAp) \leftrightarrow (\sim(pP{\sim}p) \,\&\, \sim({\sim}pPp)))$.

Hence, all the axioms of Chisholm's system are either axioms or theorems of the system we have been constructing. (Al) corresponds to (t1), (A2) corresponds to (al), (A3) corresponds to (a2), (A4) corresponds to (a3) and (A5) corresponds to (t3). Moreover, anything which can be inferred from a definition of Chisholm's system can also be inferred from either a definition, a theorem or a definition plus a theorem of our system. For (D1) corresponds to (t5) plus (d1), (D2) corresponds to (t6) plus (d4), (D3) corresponds to (d5), (D4) corresponds to (d6), (D5) corresponds to (d7) and (D6) corresponds to (t4). Therefore, every theorem in Chisholm's system is also a theorem of our system.

But not every theorem of our system is also a theorem of Chisholm's system. There is at least one immediate consequence of (t2), namely, $(\forall p)(\forall q) \sim(pUq)$, which cannot even be expressed in the vocabulary of Chisholm's systems. His logic of intrinsic value is thus properly contained in ours. And this relationship between the two systems would not be altered if we were to add to both the principle Chisholm proposes about disjunctions having mixed disjuncts.

We may conclude, then, that the objection considered in Section 2 above is not a counterexample to any of the principles of Chisholm's logic of intrinsic value. To be sure, Chisholm's system of logic cannot deal with that objection in a philosophically satisfactory fashion, for it begs the question against the objection by means of the stipulations enshrined in (D1) and (D6). However, as we have seen, the objection can be successfully met within a system which solves the problem of intrinsic comparability neither by definitional stipulation nor by theftlike postulation but by the honest toil of proof and which properly contains Chisholm's system. For this reason it seems fair to regard the system of logic presented in Section 4 as providing improved foundations for Chisholm's logic of intrinsic value.[7]

BIBLIOGRAPHY

[1] Roderick M. Chisholm, 'The Intrinsic Value in Disjunctive States of Affairs', *Noûs* 9 (1975), 295-308 [* pp. 229-39 of this volume].

[2] Roderick M. Chisholm and Ernest Sosa, 'On the Logic of "Intrinsically Better"', *American Philosophical Quarterly* 3 (1966), 244-249.

[3] Bengt Hansson, 'Fundamental Axioms for Preference Relations', *Synthese* 18 (1968), 423-442.

[4] Bengt Hansson, 'Choice Structures and Preference Relations', *Synthese* 18 (1968), 443-458.

[5] Bertrand Russell, *Introduction to Mathematical Philosophy*, (London, Allen & Unwin, 1919).

[7] I am grateful to Roderick M. Chisholm for some helpful comments.

CHAPTER 20

S. RACHELS

COUNTEREXAMPLES TO THE TRANSITIVITY OF 'BETTER THAN'

1. WHY THE THESIS IS NOT TOO RIDICULOUS TO BE TAKEN SERIOUSLY

Ethicists and economists commonly assume that if A is all things considered better than B, and B is all things considered better than C, then A is all things considered better than C. Call this principle *Transitivity.* It has great conceptual, empirical and intuitive appeal. However, I will argue that it is false.

This thesis seems ridiculous, but consider the following: 1) Some people believe that Transitivity is 'part of the meaning of' *better than*, and so Transitivity cannot be false. Without further argument, this is no more compelling than the belief that absolute simultaneity is 'part of the meaning of' *time*, and so absolute simultaneity cannot be false. 2) Our only empirical evidence for Transitivity is inductive: we have frequently observed that some A is better than some B which is better than some C which is worse than A. However, even if all the instances we have previously encountered conform to Transitivity, some unusual cases may still violate the principle. I will describe such cases below. 3) The persistent, forceful belief in Transitivity goes hand-in-hand with the assumption that value is like a line. If value can be conceived linearly, with outcomes or states of affairs represented as points along a line, then 'better than' seems transitive because 'to the right of' seems transitive. However, the idea that value can be conceived linearly is profound and difficult to justify. It is analogous to Kant's idea that time is like a line, which twentieth-century physics has brought into question. 4) Larry S. Temkin explains how three types of ethical principle, which cannot be dismissed *a priori*, threaten Transitivity: a) principles implying that in some cases different factors are relevant to comparing A to C than are relevant to comparing A to B or B to C; b) principles that are limited in scope; (c) principles implying that some morally relevant differences in degree can amount to differences in kind. I won't rehearse Temkin's explanations for why principles of types (a) and (b) threaten Transitivity, but they

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 249-263.
© 2005 *Springer. Printed in the Netherlands.*

do.[1] My counterexamples to Transitivity employ a principle of type (c): pleasures and pains enormously different in intensity differ in kind. Temkin endorses this type of counterexample, using arguments based on earlier drafts of this paper.[2]

2. FOUR PRELIMINARY POINTS

It may be useful to make the following points explicit before introducing the counterexamples:

1. Transitivity is formulated in terms of the concept *better than.* It stands or falls with the corresponding principle for *worse than.*
2. Although Transitivity refers to only three outcomes – A, B, and C – cases involving more outcomes can contradict it. It can easily be proved that a set of outcomes violates Transitivity if those outcomes can be ordered so that each subsequent outcome is worse than the one prior even though the last outcome is not worse than the first.
3. The outcomes in my examples have implicit *ceteris paribus* clauses. I will say that one outcome is 'better than' another based on incomplete descriptions; these descriptions may be amplified only such that the added details do not alter the overall value of the outcomes.
4. These examples employ stipulated, technical concepts of 'pleasure' and 'pain.' F's total conscious state is pleasurable just in case that experience, considered merely as feeling, is preferable to temporary unconsciousness; F's total conscious state is painful just in case temporary unconsciousness is preferable to that experience considered merely as feeling. Stipulating that 'pleasure' and 'pain' be understood in these ways commits us to nothing substantive and facilitates the articulation of the examples. The most important difference between these concepts of 'pleasure' and 'pain' and the ordinary ones is that mild pains, on this view, are more unpleasant than the sensations we normally call mildly painful. Ordinarily we say that a hangnail is mildly painful, but having a hangnail is not unpleasant enough for us to call it 'painful' here because it is usually not worse than temporary unconsciousness. Also, what we normally think of as 'mild pleasures' feel better than sensations barely better than temporary unconsciousness. In fact, many of the sensations we will here call 'mildly pleasurable' would ordinarily

[1] Larry S. Temkin, 'Weighted Goods: Some Questions and Comments,' *Philosophy and Public Affairs* 23 (1994), pp. 361-363, and 'A Continuum Argument for Intransitivity,' *Philosophy and Public Affairs* 25 (1996), pp. 193-194.

[2] 'Some Questions and Comments,' p. 363, and especially "A Continuum Argument for lntransitivity," sect. 4.

> be called ‘painful,’ since we prefer some unpleasantness to anesthesia. If we imagine a left-to-right continuum of sensation, with ecstasy on the far left and agony on the far right, then the main impact of our stipulated conceptions is that the range of pleasures expands to the right, reducing the range of pains. Throughout this paper we should keep an eye on these conceptions, so my arguments never gain mileage by association with ordinary usage.

3. THE FIRST COUNTEREXAMPLE: COMPARING LONG PERIODS OF PAIN

The first counterexample makes use of our technical notion of pain. Mild pains are conscious states that are barely worse than temporary unconsciousness. To focus our intuitions, here are two experiences that many people find mildly painful:

(i) Hospital patients being prepared for surgery are sometimes given the option of being sedated instead of being wheeled into the operating room and anaesthetized there. Many patients accept this offer, presumably because they believe that temporary unconsciousness is better than the anxious experience of being taken into surgery and put on the operating table.

(ii) Many people who anticipate working the entire evening on a tedious project due on the following day would prefer to snap their fingers and wake up the next morning, the project completed, rather than spend those hours working. Many people have such a preference because they consider temporary unconsciousness better than spending several hours labouring over an uninspiring task.

The first counterinstance to Transitivity consists in the outcomes A-Z. Z consists in a tremendously long period of time, each moment of which is slightly worse than temporary unconsciousness. For simplicity one might think of Z as consisting in many years of being wheeled into surgery or writing school reports researched in the *World Book Encyclopedia.* In imagining Z you should not imagine that each time you write a report, you anticipate writing millions more and all you can remember is writing millions before. For if that were the case, your life would be hellish, and each moment would be far worse than temporary unconsciousness. So, again to simplify, it may be helpful to imagine Z as a series of mildly painful experiences in which you neither anticipate nor remember similar episodes.

The first counterexample depends on the claim that A is worse than Z.[3] A is horrible agony for one year. The pains in Z and A *differ in kind* in the sense that no finite duration of Z is as bad as a year (or so) of A. To make this claim rhetorically compelling, I should now describe a method of excruciating torture that would hint at the enormity of A. (Wouldn't you rather be wheeled into the operating room repeatedly than have Satan do X, Y and Z to you for a year?) However, since torture is such an unpleasant subject, I leave this task to your imagination, if you doubt that Z is better than A.

Each outcome in the first counterexample involves a single person's experience:

A: 1 year of excruciating agony.
B: 100 years of pain slightly (or somewhat) less intense than the pain in A.
C: 10,000 years of pain slightly less intense than the pain in B.
D: 1 million years of pain slightly less intense than the pain in C.
. . .
Y: 1×10^{48} years of pain slightly less intense than the pain in X.
Z: 1×10^{50} years of pain slightly less intense than the mild pain in Y.[4]

What happens to the poor suffering souls after the year in A, the 100 years in B, and so on? I do not think it matters, if everyone in A-Z has the same fate. Two possible variations are Death and Normal Life.

Although A is worse than Z, the example creates a path from A to Z involving only changes for the worse. These changes are for the worse because increasing a pain's duration 100-fold offsets reducing its intensity slightly, or even somewhat. So the outcomes get worse until they are better, contradicting Transitivity. B is worse than A, C is worse than B, D is worse than C, ... and Z is worse than Y, yet Z is better than A.

Now I will respond to four objections. These are the most plausible rebuttals to this counterexample.

The first objection claims that this is a Sorites Paradox. Sorites Paradoxes are unsound arguments, so this argument is unsound. However, this counterexample is not of the Sorites type. Sorites arguments appeal to a series of steps each of which makes no difference to the application of a concept. For example, having one hair fewer makes no difference to whether someone is bald. In my argument, each step

---

[3] The familiar problems about total versus average utility arise in connection with single lives as well as with populations, and in asking which of two lives is better, it matters whether we appeal to total happiness or average happiness. Here I reject averaging because it leads to unacceptable results. It implies, for example, that a life consisting of ten years of ecstasy is better than a life containing that same decade of ecstasy plus twenty years of highly satisfactory (but not quite ecstatic) life.

[4] Anyone who doubts that 25 reductions in intensity could turn A's pains into Z's pains (such that B is worse than A, C is worse than B, etc.) may replace 'A-Z' with '1-50.' However, since the temporal difference between adjacent outcomes is 100-fold, the pains in adjacent outcomes should be sufficiently different in intensity that 26 outcomes are enough.

and each change makes a difference. Each outcome is worse than the one before it; each change in the intensity of a pain makes that pain slightly more intense, and thus worse.[5]

The second objection is that the fanciful nature of the example undermines its credibility. This reply might rely on either of two claims. The first is that our judgements about incredibly long, painful lives cannot be trusted. The second is that our theoretical concepts (such as *better than*) are only answerable to physically possible outcomes, and many of the lives in this example are impossible given the laws of nature.

These objections, even if sound, would not discredit the third counterexample (below), in which all the outcomes involve much shorter durations. However, they are not successful. In response to the first claim, most of the people with whom I have spoken have marked opinions about which outcome is preferable in each instance. Although the outcomes are bizarre, each comparison involves only two considerations: intensity of pain and length of time. Thus, we may have informed beliefs in each case. As for the second claim, one may live in pain for millions of years if vastly improbable quantum events keep one's nervous system (and other parts of the world) in proper states. The fact that this is wildly unlikely to happen does not tell against the example since our theoretical concepts should not rule out physical possibilities for being improbable. This is consistent with saying that, for everyday practical purposes, we might reasonably assume that our comparative assessments will be transitive.

The third objection claims that A is better than Z. According to this response, we have difficulty conceiving how much badness can accumulate, bit by bit, over $1x10^{50}$ years; a sufficient duration of mild pain is worse than a year of wrenching agony. This claim can be defended by a clever argument:

P1) One million years of mild pain are worse than three seconds of agony.

P2) One year of agony (=A) is just 10 billion three-second periods of agony strung together.

Cl) Therefore, 10 billion million years of mild pain (= 1×10^{16}) are worse than one year of agony (=A).

C2) So Z is worse than A, since 1×10^{50} years of mild pain (=Z) is even longer than the 1×10^{16} years of mild pain which is worse than A.

To analyse this argument in detail would introduce issues excellent at dividing opinion. But perhaps we can agree on one weakness. P1 claims that there is no agony so wrenching that three seconds of it are worse than a sufficiently long duration of mild pain. Is there no agony so extreme? Bear in mind, again, that the

[5] Temkin ('A Continuum Argument for Intransitivity,' sect. 5) rebuts this objection at greater length in a different way.

mild pain does not become worse over time because the person loses hope and feels demoralised; that tendency must be stipulated away or counterbalanced. Edmund Gurney says there is such agony; Henry Sidgwick says there is not.[6] The only way to investigate whether humans have experienced such pain is to interview thoughtful people who have recently suffered. Many people, I predict, would rather feel mild pain for untold aeons than their worst pains for three seconds. And many of those people will stick with this preference even if asked not to discount mild pains because they are in the farther future. If so, then we should reject the third objection. However, even if reliable parties prefer the three seconds of agony, the objection fails if physically possible pains more intense than anyone has experienced are worse than any duration of mild pain. That possibility is highly likely.

The fourth and most tempting objection is that at least one of our other judgements is incorrect. According to this response, it is better to experience some slightly less intense pain rather than a slightly more intense pain, even though the less intense pain lasts 100 times longer. This claim is tempting only for pains that are mild: one might hold that it is good for a mild pain to become more mild even if its duration increases 100-fold. However, I have not encountered anyone who believes that it is all things considered good for a severe pain to become slightly less intense and 100 times as long.

This objection ignores our stipulated conception of pain. The judgement on which it rests is plausible only if we mistakenly conceive 'mild pain' as being better than, or about as good as, temporary unconsciousness. But if we bear in mind that even the mildest pains are at every moment worse than temporary unconsciousness, then we should prefer to increase their intensity slightly if that would reduce our pain sentence by 99%. For example, we would prefer being slightly more anxious on the way to the operating room rather than repeating the slightly less anxious experience 100 times. If the objection still seems to have force, suppose it claims that V is not better than W. V is fairly mild pain for many years; W is slightly less intense pain for 100 times as long. If so, the first example still violates Transitivity if V is better than A: for then B is worse than A, C is worse than B, D is worse than C, etc., yet V is better than A. The outcomes get worse until they are better. And V is better than one year of agony because the pains in V are fairly mild.

Other objections to the first counterexample seem less plausible and coherent, so I will discuss them in a note.[7]

---

[6] Edmund Gurney, *Tertium Quid*, Vol. I (London: Kegan Paul, Trench & Co., 1887), p. 181; Henry Sidgwick, *The Methods of Ethics*, 7th edn (Indianapolis: Hackett Publishing Company, 1906; repr. 1981), pp. 123-4, fn. 1.

[7] (1) Parfit has suggested in conversation that one may preserve Transitivity by claiming that A is better than Z, yet combine this with the claim that one ought to choose Z over A when they are the only options. This suggestion is incoherent; someone who believed it could not adequately explain why one should choose Z over A, when A is better than Z, and they are the only options.

(2) Another view is that even if we should choose Z over A, Z would not be better than A if Z compares less favourably than A to some possible though unavailable alternatives. This strategy could be deployed not only against the judgement that A is worse than Z but against any judgement in the first counterexample. However, which of the endless unavailable options are relevant to comparing A and Z, if

4. THE SECOND COUNTEREXAMPLE: COMPARING LONG PERIODS OF PLEASURE

The second counterexample to Transitivity is analogous to the first. The difference is that the outcomes involve pleasure rather than pain. Z is many years of the mildest pleasures, or conscious states barely preferable to temporary unconsciousness. We may think of Z as consisting in less unpleasant versions of our mild pains, e.g., being wheeled through the hospital in somewhat better spirits or working on a task that is only somewhat burdensome. Or, to use Parfit's memorable example, we may think of Z as a life of the barest pleasures: muzak and potatoes.[8] Z is worse than A, which is 50 years of the most intoxicating joy.

A: 50 years of ecstasy.
B: 5,000 years of pleasure slightly less intense than the pleasure in A.
C: 500,000 years of pleasure slightly less intense than the pleasure in B.
D: 50 million years of pleasure slightly less intense than the pleasure in C.
. . .
Y: 5×10^{49} years of pleasure slightly less intense than the pleasure in X.
Z: 5×10^{51} years of pleasure slightly less intense than the mild pleasure in Y.

In these outcomes B is better than A, C is better than B, D is better than C, ..., and Z is better than Y, but Z is worse than A. The outcomes get better until they are worse,

---

a direct comparison of them is insufficient to assess their relative value? And why should we believe that the relevant unavailable possibilities would reverse our initial judgement that Z is better? These questions appear unanswerable. See Temkin, 'Intransitivity and the Mere Addition Paradox,' *Philosophy and Public Affairs* 16 (1987), pp. 180-183 and his discussion of the same material in 'Rethinking the Good, Moral Ideals and the Nature of Practical Reasoning,' Sect. L.

(3) Temkin entertains the view that we should individuate outcomes so finely that the identity of an outcome changes when 'it' is compared to different alternatives. Thus, our example would not contradict Transitivity because it would be wrong to say *Y is better than Z and Z is better than A*; we could only say *Y is better than Z and Z' is better than A.* However, such a view implies that Transitivity, if the principle is even coherent, applies to nothing: for it would never be the case that A is better than B and B is better than C. Furthermore, even if one argues in an *ad hoc* way that outcomes should be finely individuated only when they threaten Transitivity, such a proposal saves Transitivity only by admitting that the principle cannot always help us choose among several alternatives when our comparative preferences are clear. This all but admits that Transitivity sometimes fails. See Temkin, 'Rethinking the Good, Moral Ideals and the Nature of Practical Reasoning,' Sect. K.

(4) Another proposal for preserving the transitivity of 'better than' is mentioned in passing by Temkin: 'More cautiously, one may decide that the concept of "better than" is limited in scope, and that for many cases [i.e., apparent counterexamples to Transitivity] one needs another concept for comparing alternatives that is similar in meaning, but intransitive.' (Temkin, 'Weighted Goods: Some Questions and Comments,' p. 361, fn. 22). This suggestion is dubious both because it is *ad hoc* and because it needlessly multiplies concepts: a concept that is not transitive can be employed in all our comparative, normative judgements.

[8] Derek Parfit, 'Overpopulation and the Quality of Life,' *Applied Ethics*, (ed.) Peter Singer (Oxford: Oxford University Press, 1986), pp. 160-161.

so Transitivity is violated. Objections analogous to those against the first counterexample apply here. We need not reconsider the first and second objections (the Sorites objection and the Fanciful Nature objection) because our assessments of them would not change.

Objections three and four should be reconsidered. These objections challenge our judgements of the relative value of specific outcomes. Given that pain and pleasure may be disanalogous in various respects, these objections might prove more powerful against the second counterexample than against the first.

Objection three claims that our inability to conceive vast durations distorts our judgement: Z is better than A. We do not realise how much goodness can accumulate, bit by bit, over $5x10^{51}$ years; a sufficient duration of mild pleasure is better than fifty years of ecstasy. But I have found few people who would prefer aeons of the drabbest pleasures to a lifetime of bliss. Derek Parfit, for example, says that he would prefer 'the Century of Ecstasy' to 'the Drab Eternity.'[9] The Drab Eternity offers only the pleasures of muzak and potatoes. Though meagre, muzak and potatoes can be imagined to offer a life that is at each moment better than temporary unconsciousness.

If you accepted the third objection to the first counterexample, then you may be less tempted here. That objection claimed that a sufficient duration of mild pain is worse than a year of wrenching agony. If that claim persuaded you, then you are probably impressed by the awful nature of mild pain. And if you are so impressed, then you may be correspondingly unimpressed by the tepid nature of mild pleasure, since mild pleasures and mild pains are near neighbours on the hedonist's continuum. For that reason you may agree with me that fifty years of ecstasy are better than a lengthy duration of meagre pleasure.

The third objection, however, can be defended as follows:

> P1) One million years of mild pleasure are better than three seconds of bliss.
> P2) 50 years of ecstasy (=A) are just 500 billion three-second periods of bliss strung together.
> Cl) Therefore, 500 billion million years of mild pleasure (5×10^{17}) are better than fifty years of ecstasy (=A).
> C2) So Z is better than A, since 5×10^{17} years of mild pleasure (=Z) is even longer than 5×10^{17} years of mild pleasure which is better than A.

This argument raises controversial issues. But perhaps we can agree (in analogy to the first counterexample) that three seconds of the most intense ecstasy are better than any duration of the mildest pleasures. Dostoevsky wrote about such sensation in a letter:

---

[9] *Ibid.*, pp. 160-161. Also see James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Oxford University Press, 1986), p. 86.

> In certain moments, I experience a joy that is unthinkable under ordinary circumstances, and of which most people have no comprehension. Then I feel that I am in complete harmony with myself and the whole world, and this feeling is so bright and strong that you could give up ten years for a few seconds of that ecstasy – yes, even your whole life.[10]

The third objection fails because a sufficient difference in degree makes pleasures different in kind.

Objection four claims that it is unwise to trade one pleasure for another that is slightly less intense but 100 times as long. Again, such an objection is only plausible for the mildest pleasures, and even then only if we mistakenly suppose that some pleasures are at least as bad as temporary unconsciousness. For although we should not trade experiences of small value for longer experiences of no value, we should trade experiences of small value for experiences 100 times as long with less, but at each moment some, value.

5. THE THIRD COUNTEREXAMPLE: TWELVE BAD HEADACHES

The third counterexample involves headaches of various intensities and durations:

- A: 5 minutes: a wrenching migraine headache. Your head is ready to explode.
- B: 10 minutes: a pounding migraine headache somewhat less bad than the headache in A.
- C: 20 minutes: a hideous headache somewhat less bad than the headache in B.
- D: 40 minutes: a terrible headache somewhat less bad than the headache in C.
- E: 90 minutes: a dreadful headache somewhat less bad than the headache in D.
- F: 3 hours: a headache somewhat less bad than the headache in E.
- G: 6 hours: a headache somewhat less bad than the headache in F.
- H: 12 hours: a headache somewhat less bad than the headache in G.
- I: 1 day: a headache somewhat less bad than the headache in H.
- J: 2 days: a headache somewhat less bad than the headache in I.
- K: 4 days: a headache somewhat less bad than the headache in J.
- L: 1 week: a headache somewhat less bad than the headache in K. Its pains are only slightly worse than temporary unconsciousness.

As we move down the alphabet, the headaches get all things considered worse because having a painful headache is worse than having a headache somewhat more painful for only half as long. Transitivity is violated because A is worse than L yet it

---

[10] Quoted in Geir Kjetsaa, *Fyodor Dostoevsky: A Writer's Life,* tr. Siri Hustvedt and David McDuff (New York: Viking, 1987) p. 149.

becomes L via only changes for the worse.

To accept this counterexample we must answer Yes to two questions. First, can five minutes of a severe, agonizing migraine headache be worse than a week of mild pain?

One reason to answer No is that it is worse to lose a full week of normal life rather than five minutes. Our example, however, should not be susceptible to this objection because we want to compare the pains alone. We should stipulate that if one has a headache for less than a week, then one will spend the rest of the week unconscious, or at least having a life neither better nor worse than unconsciousness and no more productive than if one had the one-week headache.

People who have had severe migraine headaches are most qualified to judge whether a very bad migraine is worse than a week of mild pain. In conversation I have found that some such people believe that five minutes of a bad migraine headache are worse than a week of mild pain as we understand it. Thus, we should answer Yes to the first question.

Second, is it possible for a severe headache to be transformed into a mildly painful headache via eleven moderate reductions of intensity? Making a pain 'somewhat less intense' eleven times should make its intensity very much less severe and thereby bridge the gap between the extremes. Eleven steps may be more than we need. Again we should answer Yes.

If we accept these two answers, then we must deny Transitivity.

Before we move on, notice that we could easily construct an example analogous to this one involving pleasure. A, five minutes of the most heavenly sexual pleasures, is better than L, a week of lousy sex. (If Nagel is right that bad sex is better than no sex, then lousy sex should be better than temporary unconsciousness.[11]) And L could be reached from A via eleven moderate reductions in pleasure intensity, thus ensuring that B would be better than A, C better than B… In general, variations of the first three counterexamples can be formulated by altering a) the types of pain or pleasure involved, b) their durations, and c) the degree to which intensities change across outcomes.

6. THE FOURTH COUNTEREXAMPLE: PARFIT'S SECOND PARADOX

Creative work in philosophy typically arises out of existing ideas; so what one person does would have been done by someone else, given a few more years. But Derek Parfit has devised a problem so clever and original that it may be an exception to this rule. Parfit calls this 'the Second Paradox.'[12]

The Second Paradox is a set of outcomes ordered so that the outcomes seem to get better and better, yet the last outcome is worse than the first. The paradoxical conclusion, derived with Transitivity, is that the last outcome is better than the first.

---

[11] Thomas Nagel, *Mortal Questions* (Cambridge: Cambridge University Press, 1979), p. 52.

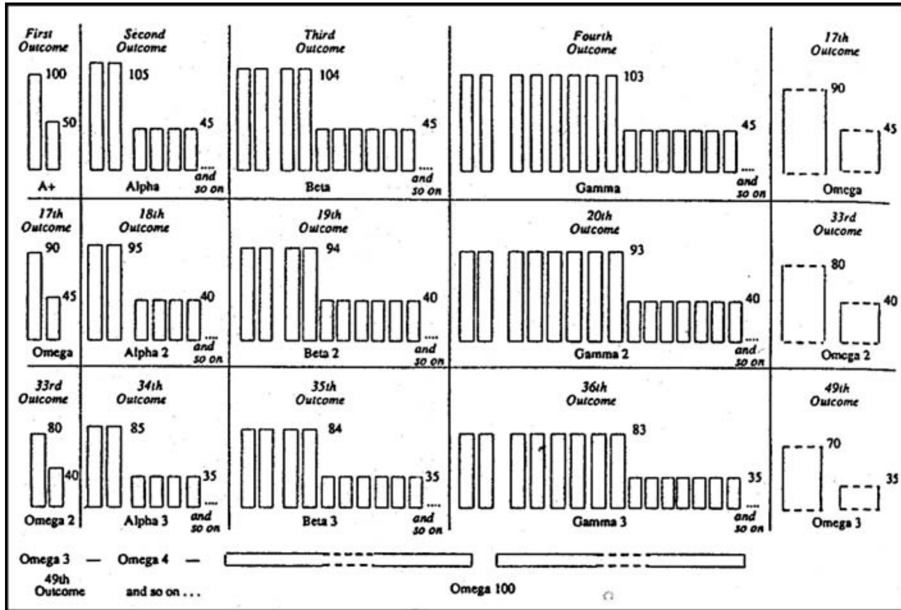
[12] Parfit discusses the Second Paradox in *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 433-437 and 'Overpopulation and the Quality of Life,' pp. 156-164. My account of the Second Paradox is taken from the latter essay in which Parfit introduces his proposal for resolving it.

Parfit does not reject Transitivity; he resolves the paradox differently. I will argue that Parfit's and Temkin's resolutions of the paradox are inadequate, and that the Second Paradox constitutes a counterexample to Transitivity.

The Outcomes in the Second Paradox

The Second Paradox is tedious to work through. However, once grasped, it dazzles the philosophical imagination.

The first outcome is A+. (See the diagram.) A+ contains two groups of 10 billion people: one group whose lives are at '100,' an ecstatic level, and another whose lives are at '50,' a level of pleasure well worth enjoying. The last outcome is Omega 100, a world that contains many, many lives each of which is barely worth living at each moment. In Omega 100 muzak and potatoes are the only pleasures in life. Although A+ is better than Omega 100, A+ is transformed into Omega 100 via only changes for the better.



The width of the blocks indicates the number of people living; the height shows the intensity of their pleasures. Dotted lines indicate that the block is much wider than shown, and wide blocks have been condensed in succeeding rows to make the diagram size manageable.

A Visual Guide to Parfit's Second Paradox

Each change from A+ to Omega 100 takes one of two forms.

The first kind of change occurs as A+ becomes Alpha. This happens by raising both groups in A+ to a 105 level of pleasure and adding many, many groups of 10

billion people whose lives, at 45, are well worth living. The outcome is much improved going from A+ to Alpha because all the people in A+ benefit from the change, especially those in the 50 group, and the only 'cost' of this benefit is adding people to the world who are glad to be alive.

The second kind of change occurs as Alpha is transformed into Beta. This occurs by lowering the two better-off groups in Alpha from 105 to 104 but raising as many worse-off groups from 45 to 104. (Even after this change there are many groups at 45.) This kind of change occurs down the Greek alphabet until we reach Omega. In Omega, there are many groups at 90 but many still at 45.

Omega is transformed into Alpha 2 by improving all the lives in Omega to 95 (including the lives that were at 45) and adding many more groups at 40. This repeats the first sort of change. Alpha 2 is transformed into Beta 2 by lowering the better-off groups to 94 but raising the same number of worse-off groups to 94. By the time we reach Omega 2, the better-off groups are down to 80, though there are many more of them, while there are still many groups at 40. At Alpha 3 all the people in Omega 2 are promoted to the level of 85 and many groups at 35 are added.

So at each Omega the average quality of life is lower than it was at the previous Omega, and the population has been greatly increased. At Omega 100, everyone's life is barely worth living at each moment. We want to say both that Omega 100 is worse than A+ and that each change from A+ to Omega 100 is for the better. Each change seems to be for the better because the quality of life is only lowered for those who are better off, and then only when this loss is more than offset by gains for the worse off.

The Second Paradox may also be formulated in terms of painful lives rather than pleasurable ones. To do this, change the numbers in the Second Paradox to negative numbers; then the outcomes will get worse until they are better (the last outcome better than the first), thus contradicting Transitivity.

Parfit's Suggestion

Parfit suggests that we resolve the paradox by claiming that Alpha is better than Beta, i.e., it is better to have 20 billion people at 105 plus many more people at 45 rather than 40 billion people at 104 plus many more people at 45 (20 billion fewer at 45 than in Alpha).

Parfit defends this by appealing to 'Perfectionism.' Perfectionism is the view that

> …even if some change brings a great net benefit to those who are affected, it is a change for the worse if it involves the loss of one of the best things in life.[13]

So Alpha is better than Beta because in Alpha the best things are better. In Alpha the luckiest 20 billion listen to Mozart; in Beta, 40 billion listen to Haydn.[14]

---

[13] Parfit, 'Overpopulation and the Quality of Life,' p. 163.
[14] *Ibid.*, p. 164.

However, we are assuming that the only factor relevant to the quality of life in the Second Paradox is the intensity of the pleasures. (We are not assuming that the 'best things in life' – Mozart's music, for example – are lost as the quality of life drops; the quality of life could fall because the best things are appreciated less.) Thus, Perfectionism is only relevant as a defence of Parfit's resolution if we interpret 'the best things in life' as the most intense pleasures. Yet on this interpretation, Perfectionism is wildly implausible. Perfectionism would imply, about our own lives, that three minutes of the best pleasure are better than fifty years of pleasure only slightly less intense. As Parfit admits, '[Perfectionism] conflicts with the preferences most of us would have about our own futures.'[15] Furthermore, Perfectionism applied to populations would imply that it could be bad for the middle class to become rich at a tiny cost to the wealthy. Hence, Perfectionism has little plausibility, and so does the claim that Alpha is better than Beta.

Temkin's Suggestion

Temkin suggests that many of the moves in the Second Paradox can be blocked by egalitarian considerations. Consider the comparison between A+ and Alpha. The fact that Alpha includes many more worse off people than A+ may represent an egalitarian respect in which Alpha is worse than A+. Perhaps this respect outweighs other factors and yields the result that A+ is all things considered better than Alpha. Or consider the comparison between Alpha and Beta. The inequality between the better off groups and the worse off groups may be worse in Beta because Beta has even more better off people that the worse off can resent (or that we can resent on their behalf). And again, perhaps this consideration suffices to license the judgement that Beta is all things considered worse than Alpha.[16]

Temkin's proposals depend on the claim that inequality is intrinsically bad. Temkin never argues this, but in *Inequality* he offers the "...fundamental intuition underlying egalitarianism – that it is bad, unfair or unjust, for some to be worse off than others through no fault of their own..."[17] Do such intuitions apply to the Second Paradox? Is it unfair, for example, that in Beta so many people prosper at the 104 level, while others only prosper at the 45 level? Nothing deters us from supposing that, in our version of the Second Paradox, the different groups of 10 billion people live in different galaxies and cannot know of each other's existence. In such a scenario, intuitions about fairness and justice do not come into play. Saying that a universe is made worse by the mere fact that isolated populations within it have varying degrees of welfare is to make an ethical judgement too much on the model of certain aesthetic judgements. It would lessen the value of a Picasso (= prospering population) to add a section onto it painted by a lesser artist (= less prosperous

[15] *Ibid.*, p. 164.

[16] Temkin suggested this in correspondence, but his published work also bears on these issues, especially *Inequality* (Oxford: Oxford University Press, 1993), chs. 7 and 9.

[17] Temkin, *Inequality*, p. 290.

population). But it would no more lessen the world's value that a remote population prospers more than we earthlings than that a possible but non-actual population prospers more than we do. Egalitarian considerations should not affect how we evaluate the Second Paradox. At the very least, they do not outweigh benefits enjoyed by billions of people. There is a treatment of the paradox more palatable than this.

A Different Proposal

We should resolve the Second Paradox by reassessing our concepts rather than changing our ethical judgements. If we deny Transitivity, then we may hold that the outcomes in the Second Paradox become better and better *and* Omega 100 is worse than A+. The Second Paradox, like the first three counterexamples involves extreme pleasures differing in kind and constitutes a counterexample to Transitivity.

7. RATIONALITY

If our examples succeed, then it is sometimes rational to prefer (in isolation) X to Y, Y to Z, and Z to X. Robert Nozick explains the 'money pump' objection to the idea that such intransitive preferences are rational:

> The idea is that with non-transitive preferences, for example preferring x to y, y to z, and z to x, a person who starts with z can be led to pay a small amount to improve his situation to the y which he prefers to it, another small amount to improve his situation to the x which he prefers to y, and then another small amount to improve his situation to z – the very z he started with – which he prefers to x, thus ending up a net loser. The argument assumes that a person is always willing to act on each individual preference, considered in isolation, and willing to act on each one repeatedly, no matter what he may know about how all of them hang together, no matter how he may foresee his sequential action on individual preferences leading him into just this sort of trouble. This is certainly an implausible assumption.[18]

It is an implausible assumption; we should reject it. A rational person with intransitive preferences and adequate information cannot be money-pumped because she will reject the principle, *it is always wise to give up something to get something better.* By denying this principle she may avoid both contradiction and poverty. For example, having Z, she may consistently (and wisely) refuse to embrace Y, even though she knows that Y is better than Z.

Objection: she could also embrace Y, but then decline X; or accept X and then stay put. How can we (possibly) determine what she should do?

A choice is most rational if the reasons for choosing it are stronger than the reasons for choosing any alternative. For cases that violate Transitivity, we lack a powerful reason to justify any particular choice: we cannot say *this choice yields an*

[18] Robert Nozick, *The Nature of Rationality* (Princeton: Princeton University Press, 1993), p. 140.

outcome that is best. If we cannot find an acceptable reason to prefer any choice to any other, then we may conclude that all available *choices* are equally wise. (This does not imply that all available *outcomes* are equally good.) If so, then a rational person could not be pumped of a cent: she would keep what she originally has (whether X, Y or Z), not parting with a penny for something better.

However, there may be sound principles that justify preferring some choices to others in cases that violate Transitivity. These principles would be sensitive to the nature of the particular outcomes involved. We will not discuss what those principles might be. Suffice it to say, if such principles imply that one outcome among several is most rationally targeted, then a rational person could be pumped of money only until she lands on that choice. She would not go around the bend.

So, I see no reason to accept Parfit's view that the failure of Transitivity implies scepticism about practical reasoning.[19]

One final remark about moral reasoning. Abandoning Transitivity has few practical implications. Because there are few exceptions to the principle, it may persist in our reasoning as a rule of thumb. If we wish to avoid using rules of thumb, then we must make fewer inferential judgements. Instead of inferring that A is better than C (given that A is better than B and B is better than C), we must compare A and C directly.

8. VALUE

The rejection of Transitivity implies that we should not conceive value quantitatively or linearly, except perhaps for some practical purposes. In this respect ethics should not strive to be scientific. Ethics is not, ultimately, a quantitative study. Even those persuaded that Transitivity is false may have trouble giving up these modes of thought. The quantitative conception of value pervades ethical theory: consider expressions such as 'the principle's weight' and 'average utility.' Even the expressions 'more value' and 'less value' connote quantities of value that are measurable and may be assigned positions on a number line. But we should not understand 'A is better than B' in terms of A having more goodness than B.

---

[19] As reported by Temkin, 'A Continuum Argument for Intransitivity,' p. 209.

CHAPTER 21

K. BINMORE AND A. VOORHOEVE

DEFENDING TRANSITIVITY AGAINST ZENO'S PARADOX[*]

Recently, Larry Temkin and Stuart Rachels have offered a new argument against the transitivity of the relationship 'all things considered better than'.[1] This argument, in its various guises, invokes our intuitions about our preferences over different bundles of pleasurable or painful experiences with different durations, which, it is argued, will typically be intransitive. This article defends against this argument the orthodox view that the relation 'all things considered better than' should be regarded as transitive by showing that Temkin and Rachels are mistaken in supposing that a preference relation satisfying their assumptions must be intransitive. It makes clear where the argument goes wrong by showing that it is a version of Zeno's paradox of Achilles and the tortoise.

Their argument centers around two kinds of counterexamples to the transitivity of 'all things considered better than', one involving painful and the other involving pleasurable experiences of differing duration. Since Temkin and Rachels both offer an example of painful experiences, we begin by discussing this case in Section 1. Section 2 explains why the argument fails. Section 3 explains why the case of pleasure is of the same form, and can be solved in the same way. Section 4 contains some general remarks on counterexamples to the transitivity of 'all things considered better than'.

[*] We thank Erik Carlson, Michael Otsuka, and Stuart Rachels for valuable discussions. Ken Binmore's work on this article was supported by the Economic and Social Research Council through the Centre for Economic Learning and Social Evolution at University College London. Alex Voorhoeve's work on this article was supported by a Cross-Disciplinary Scholarship of the Graduate School of University College London, by the Mind Studentship and the AHRB.

[1] See Larry Temkin, "A Continuum Argument for Intransitivity," *Philosophy* & *Public Affairs* 25 (1996): 175-210; Stuart Rachels, *A Theory of Beneficence* (unpublished undergraduate thesis, University of Oxford, 1993), "Counterexamples to the Transitivity of Better Than," *Australasian Journal of Philosophy* 76 (1998): 71-83 [* pp. 249-63 of this volume], "A Set of Solutions to Parfit's Problems," *Noûs* 35 (2001): 214-38, and "Intransitivity," in Volume II of *The Encyclopedia of Ethics* (2nd ed.), ed. Lawrence C. Becker and Charlotte Becker (London: Routledge, 2001), pp. 877-79.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 265-271.
© 2005 *Springer. Printed in the Netherlands.*

1. THE ARGUMENT

Temkin's and Rachels's pain counterexample rests on the following three claims, which they hold to be true:[2]

> *Claim 1*: For any unpleasant or 'negative' experience, no matter what the intensity and duration of that experience, it would be better to have that experience than one that was only a little less intense but that lasted much longer.[3]
>
> *Claim 2*: There is a finely distinguishable range of unpleasant or 'negative' experiences ranging in intensity from mild discomfort to extreme agony.
>
> *Claim 3*: No matter how long it must be endured, mild discomfort is preferable to extreme agony for a significant amount of time.

Temkin and Rachels believe that these claims taken together contradict the transitivity of 'all things considered better than'. They invite us to imagine first a lengthy life that contains an amount of excruciating torture of significant but relatively short duration. Let us call this combination of torture and the time it must be endured T_0. It follows from the first claim that one would prefer a life containing T_0 to an otherwise identical life that contains T_1, where T_1 represents slightly less intense torture for a much longer period of time. Claim 2 then allows us to apply the first claim repeatedly with a slightly lower intensity of torture and a longer time period to generate a chain of preferences, which Temkin and Rachels argue runs from T_0 (excruciating torture for a short period of time) at one end to T_{MILD} (a mild pain for a very long time) at the other. Temkin envisages the mild pain as a hangnail, Rachels as a slight headache. Each member of this chain is preferred to its successor. By transitivity, T_0 is preferred to T_{MILD}. But by claim 3, T_{MILD} is preferred to T_0, so that claims 1, 2, and 3 generate an intransitive preference.

Temkin and Rachels deduce far-reaching conclusions from their purported demonstration that it may be reasonable for people to have intransitive preferences. Rachels advocates abandoning what he calls "maximizing theories" for evaluating the goodness of states of affairs in favor of his own "quasi-maximizing theory" that embraces intransitivity.[4] Temkin quotes Derek Parfit to the effect that the argument leads to a "general skepticism about practical reasoning," and concludes that "whatever we say, in the end, about my arguments, they may require us to seriously

---

[2] Rachels, "Counterexamples," pp. 72-75 and pp. 78-79 [* pp. 251-54 and 257-58 of this volume], and Temkin, "Continuum," p. 179.

[3] Rachels, "Counterexamples," p. 73 [* p. 252 of this volume], mentions 100 times as long, while Temkin mentions twice as long. Nothing depends on this number.

[4] Rachels, "A Set of Solutions to Parfit's Problems."

rethink our understanding of the good, moral ideals and the nature of practical reasoning."[5] We believe no such drastic measures are required, since, as we show in the next section, claims 1, 2, and 3 are consistent with transitivity.

2. WHY THE ARGUMENT FAILS

One case of a person of whom claims 1, 2, and 3 are true, but who holds transitive preferences, suffices to show that the argument is wrong. A person who maximizes a utility function has transitive preferences. But a person who maximizes the utility function

$$u(p, t) = \frac{-pt}{(1+t)}$$

where u is utility, $p \geq 0$ is the intensity of pain, and $t \geq 0$ the length of time it must be endured, satisfies the three claims. Such Cobb-Douglas utility functions are standard in economics when representing tradeoffs between different commodities. The only differences are the negative sign (because the commodities are bads rather than goods) and the denominator, which represents the fact that for this person, an extra hour of pain is less troublesome after many days of pain than after a few hours of pain.

A person with utility function u satisfies claim 1, since for every level of pain, if we slightly decrease the intensity of pain and significantly increase the duration he must bear it, he will prefer the bundle with the slightly greater amount of pain. For example, suppose the level of pain for excruciating torture equals 10 and the duration the torture must be endured is 2. For a person with utility function u the disutility of this (pain, duration) bundle is 6.7. Unpleasant as it is for him, this person will still prefer this bundle to a (pain, duration) bundle of (9, 4), the disutility of which is 7.2. Utility function u is also compatible with claim 2, since it allows for a continuum of levels of pain. Finally, for a person with a utility function like u, claim 3 holds as well: a mild pain, no matter how long endured, will simply never generate the disutility of the pain of a significant amount of extreme torture. For suppose that the level of pain of a hangnail equals 1, and consider again subjecting our protagonist to excruciating torture of level 10 for duration 2. As we saw, the disutility of this amount of torture for someone with u is 6.7. Now, no matter how large we make t, the disutility of a hangnail can never exceed 1.[6] For this person, as for the "competent judges" invoked by Stuart Rachels and the audiences polled by

[5] Temkin, "Continuum," pp. 209-10.

[6] Note how our proposed utility function matches with Temkin, "Continuum," p. 192: "My model for this is something like the following. Torture's badness might range from 0 to 10, depending on its duration, with two years of torture being, say, a 7. A hangnail's badness might range from 0 to 1. Prolonging a hangnail increases the value of the decimal places representing its 'badness score', but the fundamental gap between 1 and 7 is never affected."

Larry Temkin, there is a fundamental gap between the badness of being tortured for a significant amount of time and the badness of enduring a hangnail. As Temkin writes:

> When I imagine having a hangnail for a very long time, it is not as if I imagine my situation getting closer and closer to being as bad as getting tortured for two years, but at such a slow rate that my imagination gives out long before I ever reach such a point. Rather, I imagine that there is a fundamental gap between the pain of being tortured for two years and the pain of the hangnail, and that gap is no closer to being bridged after 1000 years than it was after 100, or 50, or 10.[7]

Rachels and Temkin go wrong when they suppose that the chain of preferences $T_0 > T_1 > ... > T_n$ over bundles of intensity of pain and its duration that is generated by the repeated application of claims 1 and 2 will necessarily reach a bundle T_{MILD}, where the level of pain is equal to that of a hangnail or a slight headache. For a person with a utility function like *u,* the levels of pain in the chain $T_0 > T_1 > ... > T_n$ will converge to some limit larger than the mild pain of a hangnail or a slight headache.

It may be illuminating to note that this argument is a version of Zeno's paradox of Achilles and the tortoise. To recall, Zeno argued that if given a head start at the time t_0 when the race starts, the tortoise can never be caught by Achilles, even though Achilles is faster, since at the moment t_1 when Achilles is where the tortoise was when Achilles started, the tortoise will have moved on. When Achilles reaches the tortoise's new position at time t_2, the tortoise will have moved on again. Continuing this way, we construct an infinite sequence $t_1 - t_0$, $t_2 - t_1$, ... of the successive time intervals it takes Achilles to get to where the tortoise was when Achilles started each leg of his attempt to catch the tortoise. Zeno argued that the terms of this infinite sequence must sum to infinity, and hence the tortoise is never overtaken, but they actually sum to a finite number.

Figure 1 explains why the argument criticized in this article is a version of Zeno's paradox. The points T_0 and T_{MILD} are, respectively "two years of extreme pain" and "a very long time with a hangnail or a slight headache." With the utility function *u,* a person will be indifferent between all combinations of pain and duration that lie on a curve such as that drawn through T_0. The arrows show the direction of his preferences: he will prefer all points to the left of an indifference curve to any point on the curve, and he will prefer any point on the curve to all points that lie to the right of this curve. Note that each indifference curve has a vertical asymptote. The indifference curve drawn through T_0 has a vertical asymptote at the pain level *a.* As we decrease the pain intensity while remaining on the indifference curve (thereby remaining at the same utility level) the duration of the pain increases to infinity, but the pain intensity never falls below *a.*

---

[7] Temkin, "Continuum," pp. 191-92.

The points T_0, T_1, T_2, and T_3 show the first elements of the chain Temkin and Rachels must construct for their argument. Like the sum of the successive time intervals in the story of Achilles and the tortoise in Zeno's paradox, the sequence p_0, p_1, p_2, … converges to some limit ℓ. Since $\ell \geq a > p_{\mathrm{MILD}}$, p_n can never reach p_{MILD}. The attempt to construct a chain of preferences starting with T_0 and ending with T_{MILD} therefore fails, no matter how much time we associate with T_{MILD}.

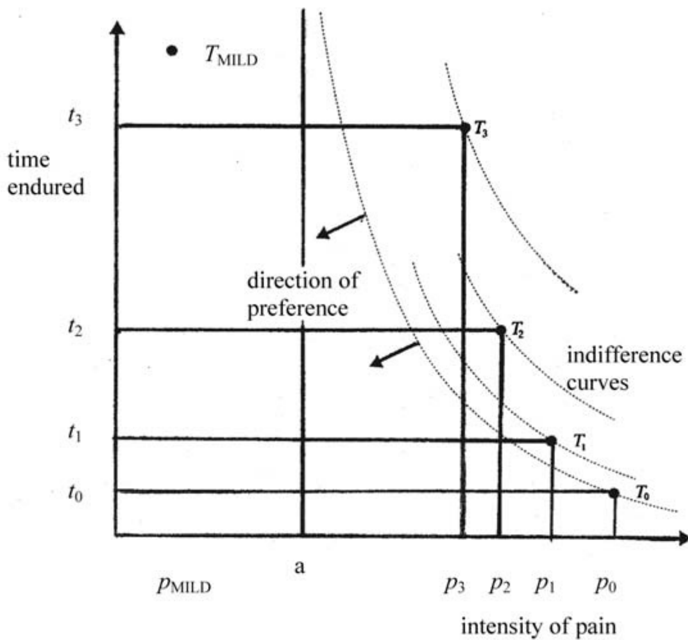


Figure 1
Indifference Curves in the Space of (Pain, Duration) Bundles

3. THE CASE OF PLEASURE

In several articles, Rachels offers an analogous counterexample to transitivity based on an analysis of a series of pleasurable experiences of differing intensity and duration.[8] He posits the following claims:

[8] Rachels, "Counterexamples," pp. 75-78 [* pp. 255-57 of this volume], "A Set of Solutions," pp. 215-16, and "Intransitivity."

*Claim 1**: For any experience of pleasure, no matter what the intensity and duration of that pleasure, it is better to experience a slightly less intense pleasure that lasts 100 times as long.[9]

*Claim 2**: There is a finely distinguishable range of pleasurable or 'positive' experiences ranging in intensity from ecstasy to mild pleasure.

*Claim 3**: No finite duration, no matter how long, of mild pleasure is as good as any significant duration of ecstasy, no matter how short.[10]

Rachels then invites us to imagine first an experience of ecstasy of significant duration. Let us call this bundle of ecstasy and time it is experienced E_0. It follows from claim 1* that one would prefer a bundle E_1, where E_1 represents slightly less pleasure lasting 100 times as long, to E_0. Claim 2* then allows us to apply the first claim repeatedly with a slightly lower intensity of pleasure and a longer time period to generate a chain of preferences, which Rachels argues runs from E_0 (ecstasy for a short, but significant period of time) at one end to E_{MILD} (mild pleasure for an extremely long time) at the other. Each subsequent member of this chain is preferred to its predecessor. By transitivity, E_{MILD} is preferred to E_0. But by claim 3*, E_0 is preferred to E_{MILD}, so that claims 1*, 2*, and 3* generate an intransitive preference.

Whether or not these claims are plausible, this counterexample fails once again because we can construct a utility function for a person of whom claims 1*, 2*, and 3* are true. A person who maximizes the utility function

$$u(e, t) = \frac{et}{(1 + t)}$$

where u is utility, $e \geq 0$ is the intensity of pleasure, and $t \geq 0$ the length of time it is enjoyed, satisfies the three claims. This is the same utility function as before but without the negative sign, because the commodities are now goods. As before, the denominator entails that an extra hour of bliss is less wonderful after many days of bliss than after a few minutes. For a person with this utility function, a night of ecstasy is worth more than a lifetime of bourgeois comfort. Indeed, it might be said to represent well the feelings of Dostoevsky, whom Rachels approvingly quotes.

> In certain moments, I experience a joy that is unthinkable under ordinary circumstances, and of which most people have no comprehension. Then I feel that I am in complete harmony with myself and the whole world, and this

[9] In "A Set of Solutions," Rachels calls this the "principle of duration."
[10] In "A Set of Solutions," Rachels calls this the "principle of lexicality."

> feeling is so bright and strong that you could give up ten years for a few seconds of that ecstasy – yes, even your whole life.[11]

We may conclude, contra Rachels, that should Dostoevsky also have held claim 1* to be true, we would nevertheless have no reason to presume his preferences were intransitive.

4. A REMARK ON COUNTEREXAMPLES TO TRANSITIVITY

Temkin's and Rachels's arguments fail. But suppose that with a different example we could all be persuaded that our choices in a pairwise comparison of three alternatives A, B, and C express a genuinely intransitive preference relation: $A > B > C >$ A. An economist would then say that this determines how we would choose from each of the feasible sets $\{A, B\}$, $\{B, C\}$, and $\{C, A\}$. He would also say, however, that we have failed to consider everything that matters because we do not specify how we would choose from the full set $\{A, B, C\}$. Whatever we plan to choose from this set, say A, we are in trouble if someone now removes C from $\{A, B, C\}$ on the grounds that its presence is irrelevant because it is not going to be chosen and because our valuation of the options A and B is independent of its presence or absence. When C is gone, we are left with $\{A, B\}$, from which we are committed to choose B. We believe that a notion of rational choice that allows such violations of the Independence of Irrelevant Alternatives would lack any useful content.[12]

---

[11] Rachels, "Counterexamples," p. 77 [* p. 257 of this volume].

[12] The Independence of Irrelevant Alternatives was introduced by John Nash in "The Bargaining Problem," *Econometrica* 18 (1950): 155-62. It differs from Kenneth Arrow's well-known condition of the same name that relates individual and social preferences. Amartya Sen, in "Internal Consistency of Choice," *Econometrica* 61 (1993): 495-521, expresses regret at the fact that Nash's and Arrow's conditions are often confused and renames Nash's condition "basic contraction consistency." If a choice function C specifies for any admissible nonempty set S of alternatives a nonempty subset $C(S)$ called the choice set of S, then Sen formulates the condition as

$$(x \in C(S) \text{ and } x \in Q \subset S) \text{ implies } x \in C(Q).$$

For the condition to make sense, the alternatives must be formulated in a manner that makes it possible to eliminate alternatives from a feasible set without altering our relative valuation of the alternatives that remain. Sen describes cases where this property appears to be violated by a rational chooser, but all his examples involve cases where the removal of one of the unchosen options changes the (expected) worth of the remaining alternatives. This means that the removed options are not irrelevant to the remaining alternatives.

CHAPTER 22

E. CARLSON

INTRANSITIVITY WITHOUT ZENO'S PARADOX

Stuart Rachels and Larry Temkin have put forward a number of alleged counterexamples to the transitivity of the relation "all things considered better than".[1] Several of these cases share a common structure.[2] A representative example by Rachels is as follows:

> Each outcome in [this] counterexample involves a single person's experience:
>
> A: 1 year of excruciating agony.
> B: 100 years of pain slightly (or somewhat) less intense than the pain in A.
> C: 10,000 years of pain slightly less intense than the pain in B.
> D: 1 million years of pain slightly less intense than the pain in C.
> . . .
> Y: 1×10^{48} years of pain slightly less intense than the pain in X.
> Z: 1×10^{50} years of pain slightly less intense than the mild pain in Y.
>
> … Although A is worse than Z, the example creates a path from A to Z involving only changes for the worse. These changes are for the worse because increasing a pain's duration 100-fold offsets reducing its intensity slightly, or even somewhat. So the outcomes get worse until they are better, contradicting Transitivity. B is worse than A, C is worse than B, D is worse than C ... and Z is worse than Y, yet Z is better than A.[3]

Temkin provides a very similar case:

> [C]ompare two lives, A and B. Suppose that both A and B are lengthy – perhaps, indeed, *very* lengthy – and that A and B are similar, except that A contains two years of excruciating torture, B four years of torture whose intensity is almost, but not quite, as bad as A's... [M]ost would judge B worse than A. Next, compare B with C, where C stands to B as B stands to A...

[1] Rachels 1998 [* pp. 249-63 of this volume] and 2001; Temkin 1987, 1996, and 1997.
[2] In particular, this is true of the first three examples in Rachels 1998, the first of which is quoted below, and of the main example in Temkin 1996, also quoted below. Examples of roughly this structure date back at least to Harrod 1936, p. 148. (Harrod does not, however, explicitly draw the conclusion that betterness is not transitive.) A seemingly different kind of example appears in Temkin 1987 and 1997 and in Rachels 2001. Those cases will not be discussed here.
[3] Rachels 1998, p. 73 [* p. 252 of this volume].

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 273-277.
© 2005 *Springer. Printed in the Netherlands.*

> [M]ost would judge C worse than B... Iterations of this reasoning imply that D would be worse than C, E worse than D, F worse than E, and so on, with the intensity of the unpleasant experiences slowly, but steadily, decreasing in each successive life. Eventually...one would be comparing two alternatives, say X and Y, such that X had an annoying hangnail for a *very* long time – perhaps thousands of years – and Y had a hangnail that was almost, though not quite, as unpleasant as X's, but that lasted twice as long... [T]ransitivity implies that A is better than Y. But surely...A is not better than Y. Specifically...no matter how long one lived, the real but mild discomfort of a hangnail throughout one's life, would be preferable to two years of excruciating torture. Correspondingly, Y would be better than A, in violation of the axiom of transitivity.[4]

These examples illustrate three general claims, which, Rachels and Temkin maintain, together entail intransitivity.[5] These claims are as follows:

Claim 1: For any unpleasant experience, no matter what its intensity and duration, it would be better to have that experience than one that was only a little less intense but lasted much longer.

Claim 2: There is a spectrum of distinguishable unpleasant experiences, ranging in intensity from extreme pain to very mild discomfort.

Claim 3: A very mild discomfort for any amount of time is preferable to extreme pain for a significant amount of time.[6]

Ken Binmore and Alex Voorhoeve claim that Rachels' and Temkin's general argument is invalid; Claims 1 to 3 do not entail intransitivity.[7] The argument is, according to Binmore and Voorhoeve, "a version of Zeno's paradox of Achilles and the tortoise".[8]

Although Binmore and Voorhoeve spot a flaw in Rachels' and Temkin's argument, this flaw is a merely technical one, which can be fixed without affecting the plausibility of the argument. Rachels and Temkin take for granted that Claims 1 and 2 together imply that there is a finite sequence of successively longer experiences, A, B, C,…, Y, Z, ranging in intensity from extreme pain to very mild discomfort, such that B is worse than A, C is worse than B, …, and Z is worse than

---

[4] Temkin 1996, p. 180. He attributes the original version of this example to Rachels.

[5] Like many other philosophers, they both speak of 'intransitive' betterness, and I shall follow this established usage. Strictly speaking, however, a relation R is intransitive only if it holds for *all* elements A, B, and C, in its domain, that if ARB and BRC, then not ARC. Of course, Rachels and Temkin do not claim that betterness is intransitive in this strict sense. What they argue is really that betterness is *non*-transitive, i.e., that there are *some* A, B, and C, such that A is better than B, and B is better than C, but A is not better than C.

[6] Temkin 1996, pp. 179 and 182ff; Rachels 2001, p. 215f.

[7] Binmore and Voorhoeve 2003, pp. 274ff. [* pp. 267 ff. of this volume].

[8] Binmore and Voorhoeve 2003, p. 272 [* p. 265 of this volume].

Y. Whether this implication holds depends on how Claim 1 is interpreted. This claim is ambiguous between the following two readings, differing with respect to the scopes of the quantifiers involved:

*Claim 1**: For any unpleasant experience A of intensity i and duration d, no matter how great i and d are, there is a small difference in intensity ε, such that it would be better to have A than an experience of intensity $j = i - \varepsilon$ and duration nd, for some number n.

*Claim 1***: There is a small difference in intensity ε, such that, for any unpleasant experience A of intensity i and duration d, no matter how great i and d are, it would be better to have A than an experience of intensity $j = i - \varepsilon$ and duration nd, for some number n.

If Claim 1 is interpreted in either of these ways, the following is a suitable formulation of Claim 2:

*Claim 2**: For any intensity level i (except the lowest level, if there is one), there is a distinguishable intensity level $j = i - \varepsilon$.

Claims 1* and 2* do not together entail that there is a sequence of the required kind. To see this, suppose that there are numbers x and y, $x > 2y$, such that x represents the intensity level of an experience of extreme pain, while y represents the intensity level of a very mild discomfort. For every positive integer $n > 1$, there might be a distinct intensity level $z = y + (1/n)x$. In this sequence, obviously, $z > y$, for any n. Thus, Claims 1* and 2* could be satisfied although the sequence never reaches an experience of intensity y. Transitivity may then hold even if Claim 3 is true.

Binmore and Voorhoeve apparently read Claim 1 as Claim 1*. There is, however, nothing in Rachels' or Temkin's discussion that excludes the alternative, Claim 1** interpretation. Concerning Temkin, there is even some slight evidence in its favour. With respect to the quoted example, he writes that "the intensity of the unpleasant experiences [is] slowly, but *steadily*, decreasing in each successive life".[9] Describing the decrease as "steady" seems to suggest that it is Claim 1**, rather than Claim 1*, that he has in mind.

In any case, we may leave exegetical questions to one side, and decide to understand Claim 1 as Claim 1**. This makes Rachels' and Temkin's argument valid. Claims 1** and 2* do together imply the existence of a sequence of the required kind. Claim 1** allows us to assume that the difference in intensity is the

[9] Temkin 1996, p. 180; my emphasis.

same between any two adjacent experiences in the sequence. This ensures that we will get from an intensity *x* to an intensity *y* experience in a finite number of steps. Although Claim 1** is logically stronger than Claim 1*, its intuitive plausibility is, as far as I can see, equally great.

There is also another problem with Binmore's and Voorhoeve's objection. Interpreting Claim 1 as Claim 1* does not block the inference to intransitivity, unless we presuppose an infinite number of distinguishable intensity levels. But it is implausible to assume that humans have sufficiently fine powers of discrimination to ensure the existence of such an infinity of levels.

Rachels and Temkin could therefore rebut the objection either by interpreting Claim 1 as Claim 1**, or by arguing that there are only finitely many distinguishable levels of intensity. Yet another option would be to simply rely on the intuitive plausibility of the assumption that there is at least one sequence of the required type. Temkin is, in fact, attracted to this move:

> [E]ven if there is reason to question one or more of [Claims 1 to 3], I think *some* version of [the quoted] counterexample is almost undeniable. I find it compelling that there is some sequence from A to Y such that A is better than B, B better than C, C better than D, and so on. But I *cannot* believe that A is better than Y.[10]

This intuitive argument will, of course, gain in strength if one is able to give a plausible example of such a sequence. Rachels and Temkin maintain that they have, indeed, provided such examples.

In conclusion, Rachels' and Temkin's argument against the transitivity of betterness cannot be dismissed as a version of Zeno's paradox. There are several responses available to this charge. Whether some version of the argument is actually sound is another matter.[11]

REFERENCES

Binmore, Ken and Alex Voorhoeve, 2003, "Defending Transitivity against Zeno's Paradox", *Philosophy and Public Affairs* 31, pp. 272-279 [* pp. 265-71 of this volume].

Carlson, Erik, 2003, "On Some Recent Examples of Intransitive Betterness", in K. Segerberg and R. Sliwinski (eds.), *Logic, Law, Morality* (Uppsala: Dept. of Philosophy), pp. 181-195.

Harrod, R. F., 1936, "Utilitarianism Revised", *Mind* 45, pp. 137-156.

Rachels, Stuart, 1998, "Counterexamples to the Transitivity of *Better Than*", *Australasian Journal of Philosophy* 76, pp. 71-83 [* pp. 249-63 of this volume].

Rachels, Stuart, 2001, "A Set of Solutions to Parfit's Problems", *Noûs* 35, pp. 214-238.

---

[10] Temkin 1996, p. 181; emphasis in the original. Also Rachels can be interpreted as employing this strategy in his 1998, where he does not explicitly rely on any general principles corresponding to Claims 1 to 3. However, such principles are introduced in his 2001, where he indicates (p. 233, n. 14) that his argument is the same as in the earlier paper.

[11] My 2003 contains two far from conclusive counterarguments.

Temkin, Larry S., 1987, "Intransitivity and the Mere Addition Paradox", *Philosophy and Public Affairs* 16, pp. 138-187.
Temkin, Larry S., 1996, "A Continuum Argument for Intransitivity", *Philosophy and Public Affairs* 25, pp. 175-210.
Temkin, Larry S. 1997, "Rethinking the Good, Moral Ideals and the Nature of Practical Reasoning", in Jonathan Dancy (ed.), *Reading Parfit* (Oxford: Blackwell), pp. 290-345.

PART V

THE COMPUTATION OF INTRINSIC VALUE

CHAPTER 23

N. FEIT

THE STRUCTURE OF HIGHER GOODS

Sometimes we say that there are greater and lesser goods, but there are different senses in which one kind of good might be said to be greater, or higher, than another one. In this paper, I present a theory of value for one of these senses. Franz Brentano had it in mind, I think, when he made the following claim: "It is quite possible for there to be a class of goods which could be increased *ad indefinitum* but without exceeding a given finite good."[1] I want to explore the consequences of the thesis that there is at least one kind of good any instance of which would be intrinsically better than any number of instances, or any amount, of some other kind of good. We may call this the 'Higher Goods Thesis', or, alternatively, we might state it as follows:

> HGT There is a kind of good X and a kind of good Y, such that necessarily, the intrinsic value of any instance of X is greater than the combined intrinsic value of all and only the instances of Y.[2]

Many philosophers have made claims about value that entail HGT. One example is W. D. Ross, who considered it to be likely "that *no* amount of pleasure is equal to any amount of virtue, that in fact virtue belongs to a higher order of value."[3] In our terminology, Ross was suggesting that virtue is a higher good and that pleasure is a lower good relative to virtue. This claim seems to me to be implausible, but it is not really my goal in this paper to attack or defend any specific claim about the identities of the higher goods. Instead, I plan to proceed as follows. I will suppose that the Higher Goods Thesis is true (although I will also discuss an argument for its truth) and go on to ask the question: what must a theory of value look like, given the

---

[1] In Franz Brentano's 1907 paper "Loving and Hating," reprinted in his *The Origin of Our Knowledge of Right and Wrong*, Roderick M. Chisholm, ed. (New York: Routledge and Kegan Paul, 1969), 158.

[2] Suppose that a pair of goods <X, Y> makes HGT true: then I shall call X a *higher good,* and I shall call Y a *lower good relative to X.*

[3] W. D. Ross, *The Right and the Good* (New York: Oxford, 1930), 150.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 281-290.
© 2005 *Springer. Printed in the Netherlands.*

truth of HGT? I will suggest that to account for the truth of HGT or the existence of higher goods, a value theory must ascribe a certain kind of internal structure to the nature of the bearers of value.

There are actually plenty of good reasons to believe in the existence of higher goods. Here I shall review one recent argument due to Noah Lemos. Lemos uses a few examples to support the view that some goods are higher than others, one of which is the following comparison of alternative possible situations.

> [I]magine that W is a possible world in which there are a million people living pleasant, virtuous lives, engaged in aesthetically and intellectually excellent activities. W is a very good world containing many different sorts of intrinsic goods: pleasure, friendship, morally and intellectually virtuous activity, the appreciation of beauty, and so forth. Now suppose that in a different possible world, W′, the sole sentient beings are wormlike creatures that we'll call "O-worms." The O-worms have no self-consciousness, no awareness of the past or future, no friendship or love for one another, and, of course, no moral, aesthetic, or intellectual life whatever. They never feel any pain, but they do feel intense pleasure for a few seconds on the sole occasion in their life cycle when they reproduce. W′ is a dull world but it does seem to be a good one, containing a vast number of instances of intense pleasure. When one reflects on these worlds, it does seem plausible to think that no matter how many instances of pleasure W′ contains, it is not intrinsically better than W.[4]

I am inclined to take this example, and many relevantly similar ones, to support an argument for HGT. If we consider possible worlds that are just like W′ except for containing more O-worms and hence more pleasure – worlds W″ , W‴ , and so on – each one will be better than the last, but none will be as good as world W. And it certainly seems that if this is the case, then some human goods are higher than the good that is instantiated in all of the O-worm-worlds, namely, sensual pleasure. So it seems that HGT must be true, even if it is difficult to say precisely which goods associated with human flourishing are the higher goods (intellectual and aesthetic pleasures, virtue, friendship, goods that accrue only to whole populations or civilizations).

One might plausibly object to Lemos's argument, however, along the following lines. Perhaps the intuitive considerations introduced by Lemos support only a weaker thesis about higher goods but not HGT. On the weaker thesis, there are kinds of goods X and Y such that some (though not all) amounts of X have more value than any amount of Y. This objection does have some force: if world W contained only 10 people, say, rather than a million, the intuition that it is better than W′ might seem to disappear, or at least its strength might seem to diminish. And if only the

---

[4] Noah Lemos, *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press, 1994), 53-54. A similar description can be found in Lemos's paper "Higher Goods and the Myth of Tithonus," *Journal of Philosophy* 90 (1993): 482-496.

weaker thesis is true, then perhaps some version of the diminishing marginal value – or asymptotic – view could account for the existence of higher goods.[5]

This objection and the issues raised by it deserve more attention than I am able to devote to them here. However, I will make a couple of claims about them. First, it really does seem to me that the sensual pleasure of the O-worms is a lower good relative to some of the goods in the people-world W (friendship, for example). There may also be *other* intuitive considerations that tell more directly in favor of the stronger thesis concerning higher goods, namely, HGT. Second, there might be goods that are not identical or reducible to the experiences of individual people or other creatures. Such goods might accrue only to entire civilizations or populations of people, in virtue of certain relations between the various individuals. These sorts of good would certainly be higher than sensual pleasure, and cases like the one described by Lemos appear to me to support their existence.

I hope to show that any view entailing the existence of higher goods can be incorporated into a principled account that yields answers to the fundamental questions of value theory. The questions include at least these three: (i) Which things are the bearers of value (i.e., what are the basic goods and evils, or bads, if the subject is nonmoral value)? (ii) How does the value of a compound situation or state of affairs depend upon the values of its simpler parts? (iii) How do the values of two compound situations, or entire possible worlds, determine that one of them is better than (or worse than, or just as good as) another? (Of course, the answer to this question is very simple if the relevant values are cardinal numbers.)

The existence of higher goods, after all, appears to tell against a quite natural (although perhaps naïve) conception of value: the totalistic view. On such a view, each basic value-bearing entity has a (finite) positive or negative value, and the value of a complex situation is obtained by summing the values of the basic, value-bearing entities out of which it is compounded. The relative value of a pair of arbitrary possible situations could then be determined simply by comparing their numerical values. However, it is quite clear that if some goods are higher than others, then there is something wrong with the general strategy of assigning a finite value to each good or bad part of a complex situation and then summing these numbers to determine the value of the entire situation. To see this, suppose that X is a higher good and Y is a lower good relative to X, and suppose that some instance of X has an arbitrary finite value n. If each instance of Y also has a finite intrinsic value, then it must be possible for there to be enough instances of Y so that their combined value exceeds n; however, this would contradict the assumption that X is a higher good and that Y is lower than X.

One might try to save the totalistic view by adopting some sort of principle of diminishing marginal value for (at least) the lower goods. With respect to the example above, this would consist in attributing progressively smaller numerical values to each new instance of Y, in such a way that the combined intrinsic value of

---

[5] See, e.g., Thomas Hurka, "Value and Population Size," *Ethics* 93 (1983): 496-507 and James Griffin, *Well-Being* (Oxford: Clarendon Press, 1986), chapter 5. I will briefly return to this view later.

all and only the instances of Y falls short of *n* (the value of our instance of X). Taken as an attempt to rescue the totalistic view, however, this strategy fails since it allows qualitatively identical instances of a given kind of good (pleasure, for instance) to differ in intrinsic value. Moreover, as an attempt to account for the kind of higher goods in question here, the strategy seems to me to place a quite implausible restriction on the lower value-range of any instance of a higher good. For example, in the abstract case of X and Y above, the asymptotic view entails that *n* could not be an arbitrarily small quantity (for if it could, it would be possible for the value of all the instances of Y to exceed that of X). But restrictions upon the value *n* should flow from the nature of good X, which *n* represents; they should not be determined by a theory about the diminishing marginal value of instances of an independent good like Y.[6]

To take a concrete example of a theory of value along totalistic lines, let's consider the classical version of Millian hedonism. The basic bearers of value are experiences of pain or pleasure, each of which has a finite intrinsic value determined perhaps by its intensity, duration, and (at least in the case of pleasures) its quality (pains will get assigned negative values). On this view, the intrinsic value of a possible situation is the sum of the values of all of the experiences of pleasure and pain that obtain in it, and one situation is at least as good as another provided that it has at least as great an intrinsic (cardinal) value.

Ad hoc epicycles aside, this naïve version of hedonism is plainly inconsistent with HGT. My aim here is to characterize the structure of higher goods in general and not to inquire into the nature of Mill's axiology. However, it will be helpful to approach the general form of a theory of value that is compatible with the existence of higher goods by way of a quick look at a revised version of Millian hedonism, a version that is itself consistent with the existence of higher goods.

Suppose that Mill's higher pleasures (the pleasures of the intellect, for example) are taken to be higher goods.[7] Suppose for simplicity that there are only two orders of pleasure and that the pleasures of mere sensation are lower goods relative to the various higher pleasures (perhaps these are the moral, intellectual, and aesthetic pleasures).

Never mind whether the claim that certain instances of pleasure are higher goods is plausible or correct; instead, consider the task of assigning values to the relevant experiences in a way that conforms to our stipulation that certain of the pleasures are higher goods. Clearly, to assign a finite cardinal number to each basic instance of pleasure will fail to give the result that some pleasures are higher goods (at least if the value is determined by the intrinsic features of an experience such as the

[6] Moreover, the asymptotic view is incompatible with certain intuitions about direct proportionality. For example, in the absence of other goods, it seems that doubling the amount of a given good thing will result in a situation with twice as much value.

[7] Curiously enough, there seems to be good reason to think that this was Mill's actual view. Mill suggests that there is a kind of pleasure such that people familiar both with it and with another (lower) kind of pleasure "would not resign it *for any quantity of the other pleasure* which their nature is capable of" (my emphasis). See chapter 2 of Mill's *Utilitarianism* (Indianapolis: Hackett Publishing Company, 1979), 8.

intensity of feeling and the duration in time, as it certainly should). And yet, on any version of hedonism, it ought to make sense to speak of the intrinsic value of an experience of pleasure or pain. If this didn't make sense, it would be difficult to imagine how we could determine whether the experience would be better or worse than another.

On the assumption that some goods are higher than others, I suggest taking the value of an experience or situation to have a certain kind of structure. In particular, if the higher and lower pleasures are the only goods, then the intrinsic value of each good thing is best taken to be an ordered pair. This will make it easy to rank experiences of pleasure and pain, and the more complex situations compounded out of them, in accord with HGT. Consider the following three-part formulation of the view currently under discussion, which we might as well call 'Hedonism with Higher Pleasures'. Let 'e' stand in for any basic episode or experience of pleasure:

> HHP1 The intrinsic value of e = the pair <x, y> such that x is the quantity of higher pleasure experienced in e, and y is the quantity of lower pleasure experienced in e.

Let's assume that the values of experiences of pain or suffering are also ordered pairs of an appropriate sort, in which the relevant numerical values are negative.

If we individuate experiences of pleasure in such a way that each one consists in the experiencing of exactly one of the two relevant kinds of pleasure, then their associated ordered pairs will contain one null value. For example, if I take pleasure to degree 5 in exercising my intellect for 5 seconds, the value of the experience might be <25, 0>; however, if a pig in its slop takes pleasure in some merely sensual experience to the same degree and for the same duration, then the value of the experience could be <0, 25>. If we allowed mixed experiences, containing both sorts of pleasure, then we would get intrinsic values given by pairs like <5, 10> and <25, 15>.

How shall we obtain the value of an entire situation – or possible world, if you wish – given the values of the various experiences of pleasure and pain that occur in it? The notion of a *pair sum* will be quite helpful here. We may use the following abbreviated definition:

> $<x, y>$ is the pair sum of $<x_1, y_1>$, $<x_2, y_2,>$, … , $<x_n, y_n>$ =df. $x = x_1 + x_2 + ... + x_n$, and $y = y_1 + y_2 + ... + y_n$.

For example, the pair sum of <25, 0>, <0, 25> and <5, 10> is the pair <30, 35>. The pairs are put in a list, and then the sum of the two columns is taken, yielding a new ordered pair.

Here is what Hedonism with Higher Pleasures ought to say about how the value of a complex state of affairs is determined by the values of its constituents.

HHP2 The intrinsic value of a situation s = the pair sum of all the intrinsic values of the experiences of pleasure and pain that occur in s.

On this view, the value of a possible situation is a certain ordered pair. How shall we determine whether such a situation is better or worse than another? Here, we cannot simply compare cardinal numbers. Reflection upon the nature of higher goods, it seems, directs us instead to decide in accord with the following principle.

HHP3 Suppose that the intrinsic value of situation s_1 is the pair $<x_1, y_1>$, and the value of s_2 is $<x_2, y_2>$. Then, s_1 is at least as intrinsically good as s_2 if and only if (i) x_1 is greater than x_2, if x_1 and x_2 are not equal; and (ii) y_1 is at least as great as y_2, if x_1 and x_2 are equal.

Intrinsic betterness can then be defined in terms of the "at least as intrinsically good as" relation: situation A is intrinsically better than situation B if and only if A is at least as good as B, but B is not at least as good as A.

Let me illustrate how HHP3 works. Again, the background assumption is that experiences of certain kinds of pleasure are higher goods and, hence, that each such experience is better than an experience or collection of experiences of merely sensual pleasure. Suppose then that we have a situation in which the various sentient inhabitants experience goods of both of these sorts: it will thus have an intrinsic value such as <500, 750>. Consider another situation in which the inhabitants – perhaps they are Lemos's O-worms – can experience only sensual pleasures: it might have a value such as <0, 3000>. Since 500 is greater than nothing, HHP3 entails that the first of these situations is the more intrinsically valuable (it is at least as good as the second situation, but the second is not at least as good as it). Notice that arbitrarily increasing the quantity of lower pleasure in the second situation will never make it better than the first. This is just what we wanted from a value theory, given the present conception of higher pleasures as higher goods.

Here are a few more consequences of Hedonism with Higher Pleasures, as constituted by HHP1 through HHP3. If the quantity of higher pleasure is the same in two situations, but one of them contains a greater amount of lower pleasure, then this one is the more valuable possibility. If, however, two situations have the same amount of higher pleasure, and also of lower pleasure, then they are tied with respect to intrinsic value (since each one is at least as good as the other).

Again, I do not claim that this version of hedonism is a plausible theory of value, and I am not suggesting that Mill actually held it; I simply maintain that the view gives results that are in accord with the assumption that kinds of pleasure divide into higher and lower goods. More important for my present purposes is the fact that the view generalizes: when structure is given to intrinsic value, the existence of higher goods – of whatever sort – poses no problem for a systematic value theory.

The general idea here is relatively simple. It also bears a certain resemblance to John Rawls's notion of a *lexical order*.[8] Identifying intrinsic values with ordered structures allows us to explain how value accrues primarily to certain basic entities (to mental states or experiences, persons or persons-at-times, states of affairs, or what have you) as well as how the value of an arbitrary complex situation is determined by the distribution of value among the relevant basic entities.

It is conceivable that something that qualifies as a higher good is also a lower good relative to something else: in other words, there might be an ordering of different kinds of good and bad. To state the general theory of the structure of higher goods, suppose that there are n such kinds, for any positive integer n. The intrinsic value of a basic instance of any such good is identified with an n-tuple, in a way analogous to the way that Hedonism with Higher Pleasures assigned values – in that case, ordered pairs – to the various experiences of pleasure. The intrinsic value of a compound situation or possible world will also be determined analogously. Each of the n-tuples that represent the basic goods and evils that occur in the situation is put on a list, and then the sum of each column in the list is obtained, the result being another n-tuple. To expand the previously introduced notion of a *pair sum,* this would make the value of such a situation the *n-sum* of the values of all of the basic goods and evils that occur in it.[9]

Finally, the general principle governing the relative intrinsic values of whole situations (or entire possible worlds) can be given as follows:

THG Suppose that the intrinsic value of situation X is the structure $<x_1,..., x_n>$, and that the intrinsic value of Y is $<y_1,..., y_n>$. Then X is at least as intrinsically good as Y if and only if:

(1) $x_1 > y_1$, if $x_1 \neq y_1$; and
(2) $x_2 > y_2$, if $x_1 = y_1$ & $x_2 \neq y_2$; and …
(n) $x_n \geq y_n$, if $x_1 = y_1$ & $x_2 = y_2$ & ... & $x_{n-1} = y_{n-1}$.

According to this theory of higher goods, there is a weak ordering of possible situations under the "at least as intrinsically good as" relation: the relation turns out to be transitive, and for any pair of situations, either the first is at least as good as the second or the second at least as good as the first. This result seems to me to be quite plausible in itself, and moreover it would seem to be required by any plausible consequentialist ethical theory.

---

[8] See John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press, 1971), especially 42-45. The analogy between a structured value and a lexical or serial order should be clear to anyone familiar with Rawls's discussion. But Rawls's notion applies to *principles* (i.e., certain principles about justice are the things that get ordered), whereas my view is concerned with the measurable values of states of affairs or other valuable entities.

[9] Perhaps we shouldn't require that *addition* be the operation performed on the columns in this list of *n*-tuples. The abstract theory of higher goods to follow is compatible with the claim that the relevant values are determined by some other function or operation.

Let's consider how this type of theory might apply to a particular example. Derek Parfit discusses a case in which he has a choice between two possible futures:

> I could live for another 100 years, all of an extremely high quality. Call this *the Century of Ecstasy.* I could instead live for ever, with a life that would always be barely worth living. Though there would be nothing bad in this life, the only good things would be muzak and potatoes. Call this *the Drab Eternity.*
>
> I believe that, of these two, the Century of Ecstasy would give me a better future. And this is the future that I would prefer.[10]

Parfit thinks that no amount of the value that would accrue to him as he lived out each day of the Drab Eternity would be as good for him as living through the Century of Ecstasy. He says, "the Century of Ecstasy would be better for me in an essentially qualitative way."[11] According to the general view presented here, there is also a more quantitative way in which the Century of Ecstasy is more valuable: this is because some of the goods associated with the Century of Ecstasy seem to be higher than all of the goods associated with Drab Eternity. So, the value of the Century of Ecstasy (for Parfit) might be something like <100, 36,500>, and the value of the Drab Eternity (for him) might be <0, 999,999>.[12] Consequently, no amount of the value provided to Parfit by the Drab Eternity would make this outcome better than the Century of Ecstacy.[13]

I would like to conclude with a few brief remarks on the consequences of higher goods for the theory of intrinsic value. Consider the claim that the intrinsic value of a compound situation is equal to the sum of the values of its basic parts.[14] Call this claim the 'Principle of Summation'.[15] Lemos asserts that the existence of higher goods is incompatible with this principle, as well as with a more general principle about quantifying intrinsic value.

---

[10] Derek Parfit, "Overpopulation and the Quality of Life," in *Applied Ethics,* Peter Singer, ed. (New York: Oxford, 1986), 160.

[11] Ibid., 161.

[12] To avoid irrelevant complications, suppose that the Drab Eternity lasts only 999,999 years.

[13] There may be goods higher than any of those associated with the Century of Ecstasy: perhaps these would be goods that can accrue only to whole populations or communities or civilizations. The point is that there is a way for a systematic theory of value to accommodate Parfit's intuitions about these two outcomes.

[14] There are difficult questions concerning the nature of the relevant part/whole relation and the parts that are to count as basic. I think these can be answered in a satisfactory way. Any claim about summation, at any rate, depends upon there being satisfactory answers to these questions.

[15] Many philosophers have held that there is some sort of numerical measure of intrinsic value. One contemporary author who defends the Principle of Summation is Fred Feldman, in *Doing the Best We Can* (Dordrecht: D. Reidel Publishing Co., 1986), 218. Bentham seems to have held a similar view. Many authors have also rejected the notion of summation (although in most cases the principle under discussion is not clearly the version I have given in the text). For example, see G. E. Moore, *Principia Ethica* (Buffalo: Prometheus Books, 1988), 27-29; Ross, *The Right and the Good*, 138; and perhaps Rawls, *A Theory of Justice,* 60-83 (especially 61 and 82-83). For a very interesting discussion of Brentano's rejection of summation, see Roderick Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986), especially 69-72 [* pp. 305-307 of this volume].

> If we accept the existence of such higher goods, then we must reject both (i) the principle of summation and (ii) the assumption that there is a numerical measure of intrinsic value whose values can be added, subtracted, multiplied, etc.[16]

Lemos is correct to point out that the existence of higher goods is incompatible with the Principle of Summation. However, if higher goods behave in the way described here, then there might very well be a straightforward substitute for the Principle of Summation. For example, Hedonism with Higher Pleasures implies the principle of *pair-summation*: that the intrinsic value of a whole is equal to the pair sum of the intrinsic values of its parts. And supposing there are *n* incommensurable kinds of good, a corresponding principle of *n-summation* might very well hold. So the initial plausibility of the Principle of Summation, before higher goods are countenanced, appears to carry right over to its substitute.

Moreover, if higher goods have something like the structure I have attributed to them here, then it is misleading at best to claim that there can be no numerical measure of intrinsic value. The view just outlined, for example, implies that intrinsic values are ordered pairs of numbers, and that arithmetical operations on the values of basic goods are essential to determining the value of an entire outcome. Clearly, someone who objected to the use of cardinal numbers in assigning values to states of affairs would have no *additional* reason to object to them, were she to come to embrace higher goods. Even if the phrase "numerical measure of intrinsic value" does not correctly apply to sequences or *n*-tuples of numbers, I hope to have shown that there is a simple and principled way to compare situations that have higher goods with those that do not.[17]

REFERENCES

Chisholm, Roderick (ed.). *The Origin of Our Knowledge of Right and Wrong.* New York: Routledge and Kegan Paul, 1969.

———. *Brentano and Intrinsic Value.* Cambridge: Cambridge University Press, 1986.

Feldman, Fred. *Doing the Best We Can.* Dordrecht: D. Reidel Publishing Company, 1986.

Hurka, Thomas. "Value and Population Size." *Ethics* 93 (1983): 496-507.

Griffin, James. *Well-Being.* Oxford: Clarendon Press, 1986.

Lemos, Noah. "Higher Goods and the Myth of Tithonus." *Journal of Philosophy* 90 (1993): 482-496.

———. *Intrinsic Value: Concept and Warrant.* Cambridge: Cambridge University Press, 1994.

[16] Lemos, "Higher Goods and the Myth of Tithonus," 496. Of course some philosophers who have rejected summation have done so for reasons that have little or nothing to do with higher goods. One such reason is Moore's doctrine of organic unities, according to which the value of an entity can be a function of the relations between its parts. See, e.g., Moore, *Principia Ethica,* 27-36. This doctrine can be avoided, however, by taking intrinsic value sometimes to attach to relational states of affairs.

[17] I would like to thank Fred Feldman and Stephen Kershnar for very valuable comments on earlier versions of this paper.

Mill, John Stuart. *Utilitarianism.* Indianapolis: Hackett Publishing Company, 1979.
Moore, G. E. *Principia Ethica.* Buffalo: Prometheus Books, 1988.
Parfit, Derek, "Overpopulation and the Quality of Life." In *Applied Ethics,* edited by Peter Singer. New York: Oxford University Press, 1986.
Rawls, John. *A Theory of Justice.* Cambridge, MA: Belknap Press, 1971.
Ross, W. D. *The Right and the Good.* New York: Oxford, 1930.

CHAPTER 24

G. ARRHENIUS

SUPERIORITY IN VALUE

1. INTRODUCTION

Let's say that A and B are two types of goods such that more of A or B is better than less. I'm going to discuss the following two ideas:

> *Strong Superiority* (roughly): Any amount of A is better than any amount of B.
>
> *Weak Superiority* (roughly): Some amount of A is better than any amount of B.

It is easy to find examples of these ideas in the literature. For example, early in the 20th century Franz Brentano claimed that "[i]t is quite possible for there to be a class of goods which could be increased *ad indefinitum* but without exceeding a given finite good".[1] Likewise, W. D. Ross asserted that "[w]ith respect to pleasure and virtue, it seems to me much more likely to be the truth that *no* amount of pleasure is equal to any amount of virtue, that in fact virtue belongs to a higher order of value, beginning at a point higher on the scale of value than pleasure ever reaches…"[2] Similar views have been proposed by, among others, Roger Crisp, Jonathan Glover, James Griffin, Rem Edwards, Noah Lemos, Derek Parfit, and John Skorupski.[3] Its lineage goes back to at least Francis Hutcheson in the early 18th century and of course John Stuart Mill in the mid 19th century.[4]

Superiority in value can be compared to

[1] Brentano (1907).
[2] Ross (1930), p. 150.
[3] See Crisp (1988), p. 188; Crisp (1992), p. 151; Glover (1977), p. 71; Griffin (1986), pp. 85-6; Edwards (1979), pp. 69-72; Lemos (1993); Parfit (1986), pp. 161-4; Skorupski (1999), pp. 94-101. See also Klint Jensen (1996), pp. 90-1; Portmore (1999); Riley (1993, 1999).
[4] Hutcheson (1755), pp. 117-18 (Hutcheson completed his treatise in 1738 but it was published posthumously in 1755 – cf. Edwards (1979), p. 71); Mill (1998), p. 56. For another early source with a racist twist, see Rashdall (1907), pp. 238-9. Newman (1885), vol. I, p. 204, makes a similar claim about pain and sin.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 291-304.
© 2005 *Springer. Printed in the Netherlands.*

The Archimedean Property of Value (roughly): For any amount of A there is some amount of B which is better.

This is like the Archimedean property of the real numbers: for any positive numbers x and y, there is a natural number n such that nx is greater than y. The Archimedean property seems to capture the way we usually think about the aggregation of goods. Let's say that you are considering two holiday packages. The first is a week in Stockholm, the other a week in Copenhagen. They both give the opportunity for more or less the same activities: going to museums, city-walks in pleasant surroundings, interesting culinary experiences, and the like. You have a slight preference for Stockholm. It is possible, however, to better the Copenhagen-package by adding some extra days. It seems plausible that there is such a bettering, other things being equal, that would reverse your preference in such a way that you would prefer the Copenhagen holiday. One might think that this feature is a general property of goods, that all types of goods satisfy the Archimedean Property of Value. Strong and Weak Superiority are a denial of this claim; they are both versions of what I call *non-Archimedeanism*.

Superiority in value and non-Archimedeanism are structural features that can be true of any kind of good. I shall discuss some different ways in which these ideas can be applied to the aggregation of welfare. It is important to separate these different applications of the superiority idea, since they will yield quite distinctive views with varying intuitive support. I shall then try to give an exact and generalised statement of Strong and Weak Superiority. Using these formulations, I shall prove a general result that can be used as an argument against the existence of superiority in value in certain contexts: roughly, if one holds that some type of good A is strongly or weakly superior to another type of good B, then one is committed to holding that there are two types of goods C and D such that C is weakly superior to D although goods of type C are only marginally better than goods of type D. First, however, I shall describe the kind of problems that have motivated the recent interest in non-Archimedeanism.

2. INTRAPERSONAL AND INTERPERSONAL REPUGNANT CONCLUSIONS

During the last twenty years or so, non-Archimedeanism has again become popular in connection with theories of welfare and population ethics, especially among Oxford philosophers. What is the reason behind this return to non-Archimedeanism? Derek Parfit has brought attention to a problem for maximising theories of beneficence, such as Total Utilitarianism, which tells us to maximise the welfare in the world. They imply what he calls the Repugnant Conclusion:

The Repugnant Conclusion: For any perfectly equal population with very high positive welfare, there is a population with very low positive welfare which is better, other things being equal.[5]

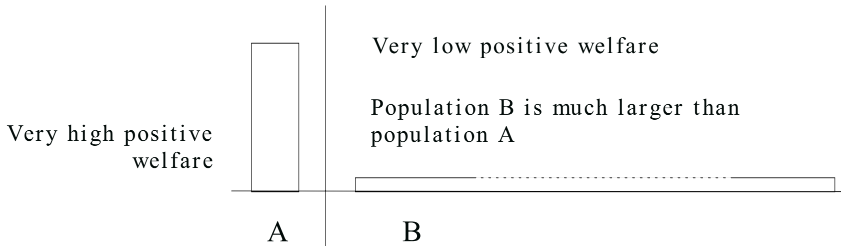


Diagram 1

In diagram 1, the width of each block represents the number of people whereas the height represents their lifetime welfare. All the lives in the diagram have positive welfare, or, as we also could put it, all the people have lives worth living. People's welfare is much lower in B than in A, since the A-people have very high welfare whereas the B-people have very low positive welfare. The reason for the very low positive welfare in the B-lives could be, to paraphrase Parfit, that there are only enough ecstasies to just outweigh the agonies or that the good things in life are of uniformly poor quality, e.g., working at an assembly line, eating potatoes and listening to Muzak.[6] However, since there are many more people in B, the total sum of welfare in B is greater than in A. Hence, Total Utilitarianism ranks B as better than A – an example of the Repugnant Conclusion.[7]

One way of avoiding this implication but retaining the idea of maximisation is to invoke a form of non-Archimedeanism and claim that there is some population of lives with very high welfare that is better, or has greater aggregate welfare, than any population of lives with very low positive welfare, although the addition of more

---

[5] See Parfit (1984), p. 388. My formulation is more general than Parfit's except that he doesn't demand that the people with very high welfare are equally well off. Although it is through Parfit's writings that this implication of Total Utilitarianism has become widely discussed, it was already noted by Henry Sidgwick (1907), p. 415, before the turn of the century. For other early sources of the Repugnant Conclusion, see Broad (1930), pp. 249-50, McTaggart (1927), pp. 452-3, and Narveson (1967).

[6] See Parfit (1984), p. 388 and Parfit (1986), p. 148. In Arrhenius (2000), I discuss different interpretations of the Repugnant Conclusion in some detail.

[7] Notice that problems like this are not just problems for utilitarians or those committed to welfarism, the view that welfare is the only value that matters from the moral point of view, since we can assume that the other things are roughly equal. We can assume that other values and considerations are not decisive for the choice between populations A and B. This is a problem for all moral theories which hold that welfare at least matters when all other things are equal, which arguably is a minimal adequacy condition for any moral theory.

lives with very low positive welfare always makes a population better, or always increases the aggregate welfare.

Diagram 1 could also be taken to represent an intrapersonal version of the Repugnant Conclusion. The width of each block would then represent the length of a life, and the height would represent the well-being at a certain time. For example, block A could represent a life of a hundred years in which every year is of very high quality, and block B could represent a very long life but in which every year is of very low quality.[8] If the B-life is just long enough, then the total welfare of that life will be greater than the A-life and thus considered better by a maximising theory of welfare. Again, this implication could be blocked by invoking some form of non-Archimedeanism. This is what Parfit suggests:

> I could live for another 100 years, all of an extremely high quality. Call this *the Century of Ecstasy*. I could instead live for ever, with a life that would always be barely worth living … the only good things would be muzak and potatoes. Call this *the Drab Eternity*. … I claim that, though each day of the Drab Eternity would be worth living, the Century of Ecstasy would give me a better life. … Though each day of the Drab Eternity would have some value for me, *no* amount of this value could be as good for me as the Century of Ecstasy.[9]

Likewise, in his influential discussion of intrapersonal aggregation of welfare, James Griffin has proposed that there can be what he calls "discontinuity" among prudential values (welfare) of the form "enough of A outranks any amount of B". Discontinuity entails, he explains:

> … the suspension of addition; … we have a positive value that, no matter how often a certain amount is added to itself, cannot become greater than another positive value, and cannot, not because with piling up we get diminishing value or even disvalue …, but because they are the sort of value that, even remaining constant, cannot add up to some other value. … [I]t is more plausible that, say, fifty years at a very high level of well-being – say, the level which makes possible satisfying personal relations, some understanding of what makes life worth while, appreciation of great beauty, the chance to accomplish something with one's life – outranks any number of years at the level just barely living – say, the level at which none of the former values are possible and one is left with just enough surplus of simple pleasure over pain to go on with it.[10]

I take it that most people, like me, find Griffin's and Parfit's examples convincing. To establish whether the non-Archimedean idea involved in the

[8] There are other versions of the interpersonal case in which the compared lives may be equally long but one life contains some amount of A and the other an arbitrarily large amount of B, where some of the B-goods come more or less simultaneously. An example of the latter case, which I owe to Wlodek Rabinowicz, is appreciation from enthusiastic readers.

[9] Parfit, (1986), pp. 161-4, emphasis in the original.

[10] Griffin (1986), pp. 85-6.

examples is plausible, however, we need to spell it out in more detail and distinguish among different varieties of it.

3. AGGREGATION OF WELFARE AND NON-ARCHIMEDEANISM

Let's call those things that are good or bad for people welfare components. Here are some components that have been proposed in the literature: pleasure or pain; taking pleasure or displeasure in worthy or unworthy states of affairs; satisfied or frustrated desires; being more or less autonomous; achieving greater or lesser accomplishments; true or false beliefs; having satisfying or dissatisfying personal relationships; experiences of beautiful or ugly objects, and so forth.

The kind of non-Archimedeanism expressed in the passages above from Griffin and Parfit is about the relationship between orderings of welfare components and the ordering of lives in respect to their welfare. However, the route from welfare components to the welfare of lives can take different paths and, which is important, some form of non-Archimedean property can appear along different stops of this path. For example, we could order experiences of pleasure, including collections of pleasurable experiences, in terms of how pleasurable they are (how much pleasure these experiences contain taken as a whole), i.e., by the relation "is at least as pleasurable (painful) as". We could then ask how such an ordering relates to the welfare of a life. We could, for example, ask whether it is always *better for* a person to experience more happiness rather than less, other things being equal, and thus whether for any amount of intense pleasure, there is always some amount of minor pleasure which is better for a person. Here we are asking whether some kind of non-Archimedean property appears in the contributive value of pleasure to the welfare of a life. On this view, it is possible that the ordering of pleasures by the relation "is at least as pleasurable" fulfils the Archimedean property.

We could also start by looking at *atomic* experiences of pleasures, understood as the shortest possible experiences of pleasure that are the building blocks of all other experiences of pleasure. We could then investigate how the ordering of the atomic experiences of pleasure relates to the ordering of all experiences of pleasures, including collections of pleasurable experiences, in terms of the relation "is at least as pleasurable as". We could then ask whether for any number of intense pleasures, there is always some number of minor pleasures which, taken together, are *more pleasurable*. Here we are asking whether the non-Archimedean property appears in the aggregation of single pleasures to the overall pleasure in a life. How pleasure contributes to well-being would then be a further question.

If the second approach is combined with the view "the more happiness, the better a life", then it might be extensionally equivalent to the first. Still, these two versions of non-Archimedeanism are different in an important way: if we hold that the non-Archimedean property appears in the ordering of pleasurable experiences by the relation "is at least as pleasurable as", then it's still possible for us to claim that "the

more happiness, the better a life". However, if we hold that the non-Archimedean property appears only in the ordering of lives by the relation "has at least as high welfare as", i.e., in the contributive value of pleasure to the welfare of a life, then we cannot claim that "the more happiness, the better a life". This might be of importance for a hedonist who thinks that this credo is an integral part of hedonism as a theory of welfare. Perhaps Mill had something like this in mind when he wrote:

> It is quite compatible with the principle of utility to recognise the fact, that some *kinds* of pleasure are more desirable and valuable than others. … Of two pleasures, if … one of the two is, by those who are competently acquainted with both, placed so far above the other that they … would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality, so far outweighing quantity as to render it, in comparison, of small account.[11]

The non-Archimedean property can also appear on a third level. We can order populations of lives, which can be outcomes of actions, in terms of how morally good they are, that is, by the relation "is at least as good as". We can ask how pleasure or welfare contributes to this value. Assume that we have an ordering of pleasurable experiences in terms of "is at least as pleasurable as". We could then ask whether for any number of lives experiencing intense pleasure, there is a number of lives experiencing only slight pleasures, which is better. We could also ask whether for any number of lives with very high welfare, there is a number of lives with slight welfare which is better. Here we are asking whether the non-Archimedean property appears in the contributive value of pleasure or welfare to the value of populations.

Of course, the non-Archimedean property could also apply to the other kinds of welfare components mentioned above. Notably, it could hold between welfare components of different kinds. For example, one could believe that no amount of slight pleasure can outweigh the loss of one's autonomy. This could be a view about how pleasure and autonomy contribute to well-being or to the intrinsic value of a life. Moreover, the non-Archimedean property might appear in the aggregation of welfarist and non-welfarist goods into a measure of the intrinsic value of actions and people. This seems to be what Ross had in mind in the quotation in the introduction and what Hutcheson had in mind when he wrote:

> As to pleasures of the same kind, 'tis manifest their values are in joint proportion of their intenseness and duration. … In comparing pleasures of different kinds, the value is as the duration and dignity of the kind jointly. … No intenseness or duration of any external sensation gives it a dignity or worth equal to that of the improvement of the soul by knowledge, or the ingenious arts; and much less is it equal to that of virtuous affections and actions. … By this intimate feeling of dignity, enjoyment and exercises of some kind, tho' not of the highest degree of those kinds, are incomparably more excellent and beatifick than the most intense and lasting enjoyment of

[11] Mill (1998), p. 56.

> the lower kinds. … The exercise of virtue for a short period ... is of incomparably greater value than the most lasting sensual pleasure.[12]

Let me summarise this section. The non-Archimedean property can appear at different stages in the path from goods to welfare and the value of people, populations, worlds, and actions. It can appear in the contributive value of atomic experiences of pleasure to the pleasantness of aggregates of such experiences. It can appear in the contributive value of pleasure to the welfare of a life. It can appear in the contributive value of the welfare of lives to the value of populations and worlds. It can appear in the aggregation of welfarist and non-welfarist goods into a measure of the value of actions and people. This list doesn't exhaust all the possibilities but represents the most interesting ones. Notice that these views are logically independent. For example, one could hold that slight pleasures can always outweigh intense pleasures when it comes to the welfare of a life but not when it comes to the contributive value of a life to the value of a population. For example, one might think that one life with some amount of high quality pleasures has higher contributive value to a population than any number of lives with any amount of low quality pleasures, although the latter lives may enjoy higher welfare than the former life.

4. THE GENERAL STRUCTURE OF SUPERIORITY IN VALUE

Assume that goods can be partitioned into different *kinds* defined by some predicate, e.g., pleasurable experiences, knowledge, autonomy, accomplishments, personal relationships, bodily pleasures, intellectual pleasures, being of a virtuous character, the exercise of virtue, and so forth. This gives us different domains of valuable objects. Assume that goods of the same kind are weakly ordered (reflexive, transitive, and complete) by some relation R. R could be the natural comparative relation "_ is at least as *N* as _" of the objects in question (e.g., "is at least as pleasurable as", "is at least as autonomous as", "is at least as much knowledge as", etc.) or some value relation "_ is at least as *V* as _" (e.g., "is at least as good as", "is at least as good as for a person"). We shall suppose that for any object *e* in the domain and any number *m*, the domain contains a whole composed of *m* "*e*-objects", that is, a whole composed of *m* components equally as R as *e*. For example, let's say that we are considering a certain kind of pleasure and the relation "is at least as pleasurable as". If some pleasure *e* is in the domain, so is the whole consisting of, say, ten *e*-pleasures, i.e., ten episodes of pleasures, each as pleasurable as pleasure *e*.

We shall also assume that all of the objects in domains we are discussing are positively valuable in the following sense:

[12] Hutcheson (1755), pp. 117-18.

Weak Positive Response: For any object e and any number n, $n+1$ e-objects is at least as R as n e-objects.[13]

We shall say, somewhat contrary to common language use, that a *type* of good g is a representative of an equivalence class in a certain domain. For example, let's say that the kind of good in question is pleasure and this domain is ordered by the relation "is at least as pleasurable as", and that e is an episode of pleasure. Then the representative of the set of all pleasures equally as pleasurable as e is a type of good in this domain. The expression "n g-objects" is short for "a whole composed of n objects equally as R as g". We can now state Strong and Weak Superiority as follows:

Strong Superiority: A type of good g_1 is strongly superior to another type of good g_2 iff for any n, one g_1-object is more R than n g_2-objects.

Weak Superiority: A type of good g_1 is weakly superior to another type of good g_2 iff there is a number n such that for any number m, n g_1-objects are more R than m g_2-objects.

The statements above of Weak Positive Response, Strong and Weak Superiority, are generalised to cover any kinds and types of goods and relations, including the ones discussed in section 3. We could also use the above definitions to state relations between different kinds of valuable objects or other partitions of these objects. For example, a kind of good, say knowledge, is strongly superior to another kind of good, say, pleasurable experiences, iff all types of the first kind are strongly superior to all types of the second kind.

For Griffin and Parfit, as we pointed out above, Superiority is about the relationship between orderings of welfare components and the ordering of lives in respect to their welfare. Here's how we could adapt the definitions above for that context:

Weak Positive Response for Welfare Components: For any welfare component e and any number n, a life with $n+1$ e-components has at least as high welfare as a life with n e-components, given that these lives don't involve any other welfare components.

Strong Superiority between Welfare Components: A type of welfare component g_1 is strongly superior to another type g_2 iff for any n, a life with one g_1-component has higher welfare than a life with n g_2-

[13] As Howard Sobel pointed out to me, this condition rules out a number of goods, for example, the goods of eating potatoes and listening to Muzak, since it is plausible that for many goods and most people, there can be too much of a good thing – just consider eating 10 kilos of potatoes in a day. Notice, however, that in Parfit's example above, the goods considered were not that of eating potatoes and listening to Muzak, but days with very low but positive well-being ("… each day of the Drab Eternity would have some value for me…"). The consumption of potatoes and Muzak is just the source of the well-being.

components, given that these lives don't involve any other welfare components.

Weak Superiority between Welfare Components: A type of welfare component g_1 is weakly superior to another type g_2 iff there is a number n such that for any number m, a life with n g_1-components has higher welfare than a life with m g_2-components, given that these lives don't involve any other welfare components.

We have included the clause "given that these lives don't involve any other welfare components" to make clear that these definitions only cover lives that involve welfare components of types g_1 and g_2. Weak Superiority corresponds pretty well with the idea expressed in the quotations from Griffin and Parfit above and to Griffin's "discontinuity in value" (enough of A outranks any amount of B). Strong Superiority is similar to what Griffin calls "Trumping" which "… takes the form: any amount of A, no matter how small, is more valuable than any amount of B, no matter how large."[14]

5. THE SEQUENCE ARGUMENT

I take it that it is clear that Strong Superiority implies Weak Superiority, i.e., if g_1 is strongly superior to g_2, then g_1 is weakly superior to g_2. That the following relations hold might not be equally obvious:

C2: Any descending finite sequence $g_1, \ldots, g_n$ (i.e., a sequence such that g_1 is more R than g_2, which is more R than g_3, …, which is more R than g_n), in which the first element is strongly superior to the last element, must contain an element that is weakly superior to the element that immediately follows.

C3. Any descending finite sequence $g_1, \ldots, g_n$, in which the first element is weakly superior to the last element, must contain an element that is weakly superior to the element that immediately follows.

What does this mean? Let's illustrate the implications of C3 using Parfit's example. Assume that there are days of different qualities and that these can be arranged in a descending sequence of goodness or of how much they would contribute to the well-being of a life. It seems plausible that there can be such a sequence where the difference in quality of any two adjacent days in the sequence is marginal. For example, consider two days of a life that only differ in respect to one

[14] Griffin (1986), pp. 83, 85.

pin-prick in the left thumb. Assume now, as Parfit suggests, that there are days of such quality that some number of these are better than any number of days eating potatoes and listening to Muzak. Then there are, according to C3, two types of days, call them A-days and B-days, such that some number of A-days is better than any number of B-days although the difference in quality between these days is marginal.

Another way to put it is this. If the elements in a descending finite sequence are chosen in such a way that each consecutive element is only *marginally* worse than the immediately preceding one, then it would seem that no element will be weakly superior to the element that comes next. But then, according to C3, the first element will not be weakly superior to the last one either, however long such a sequence may be. One might find this counterintuitive, since one might think that in a sufficiently long series of small worsenings one should sooner or later reach an element that is radically worse than the point of departure.

Compare with C1:

> C1: There could exist a finite descending sequence $g_1, \ldots, g_n$ in which the first element is strongly superior to the last element, but in which no element is strongly superior to the element that immediately follows.[15]

Here's a simple example to show that C1 is true. Let's say that we have a sequence that consists of three elements, g_1, g_2, and g_3. For each of these types of goods, there is a value limit l_i such that for any n, l_i is more R than $n\ g_i$, but for any other good x such that l_i is more R than x, there is an m such that $m\ g_i$ is more R than x. Moreover, l_2 is more R than g_1, and l_3 is more R than g_2, but g_1 is more R than l_3. Since l_2 is more R than g_1, and l_3 is more R than g_2, it follows that there is an n such that $n\ g_2$ is more R than g_1, and that there is an m such that $m\ g_3$ is more R than g_2. Consequently, g_1 is not strongly superior to g_2, and g_2 is not strongly superior to g_3. However, since g_1 is more R than l_3, it follows that for any n, g_1 is more R than $n\ g_3$. Thus, g_1 is strongly superior to g_3. In other words, the first element in the sequence is strongly superior to the last element, even though no element is strongly superior to the one that comes next. So C1 is true.[16]

---

[15] Ryberg (2002), p. 418, claims that "[if] there is a discontinuity between the values … at each end of the continuum, then at some point discontinuity must set in". Rabinowicz (2003) denies this claim. As can be seen from C1-C3 above, it all depends on what one means by "discontinuity" and Ryberg isn't very clear on this point. If by "discontinuity" one means "strong superiority" then Ryberg's claim is false, as C1 shows. If one means "strong or weak superiority", then Ryberg's claim is true, as C2 and C3 show. Rabinowicz is right, however, in pointing out that Ryberg's reasoning is based on the fallacy of identifying superiority with "infinite betterness". For a detailed discussion of Ryberg's argument, see Rabinowicz (2003) and Arrhenius and Rabinowicz (2003).

[16] Here's a numerical version of the same example which I owe to Wlodek Rabinowicz. Let's say that the numerical representation of the values of g_1, g_2, and g_3, as measured on a ratio scale, are 5, 3, and 2. Suppose that the contributive value of an extra object is half of the contributive value of the preceding object of the same type. Thus, the value of one g_3-object equals 2, the value of two such objects equals 2 + 1, the value of three such objects equals 2 + 1 + ½, and so forth. It follows that there is a value limit that

I take C2 and C3 to be a quite powerful argument against the existence of superior goods in contexts where one can construct a sequence in which the value difference between adjacent goods is marginal. As I have indicated above, this is plausible in the context of well-being and I think it probably holds true for most, if not all, kinds of goods. This can only be determined on a case-by-case basis, however, and I shall not undertake that investigation here. At any rate, in such cases I find it counterintuitive that a sufficient number of objects of one type can make a whole that is better (better for, more pleasurable, etc.) than *any* whole, however large, that consists of just slightly worse objects and in spite of this whole getting better for each additional object.

One could, however, also argue that the lesson we should draw is that there is an important difference between Strong and Weak Superiority.[17] The former is a rather dramatic form of superiority in value, but the latter is not. If an object g_1 is only marginally worse than another object g_2, then g_1 will not be strongly superior to g_2. But, contrary to what one might have expected, it may still be the case that g_1 is weakly superior to g_2, even though the value difference between the two objects is marginal.

Although I grant that this is a possible interpretation of the result, I'm not sure how it would eliminate the counterintuitive result of Weak Superiority pointed out above. Moreover, Strong and Weak Superiority share a number of problems that occur when we consider mixed outcomes, that is, outcomes that involve both superior and inferior goods. Here's an example involving intrapersonal aggregation of welfare. Consider three welfare components g_1, g_2, and g_3 such that g_2 is only marginally worse than g_1 whereas g_3 is clearly worse than g_2 but still good. Assume further that g_1 is strongly or weakly superior to g_2. Consider a life *a* with a sufficient number *n* of g_1-components as compared to a life *b* with the same number of g_1-components and a much greater number *m* of g_3-components. Since the only difference between these lives is that life *b* has an additional great number of the things that make a life better, it seems reasonable to say that life *b* has higher welfare than life *a*. Consider now a life *c* with $n+m$ g_2-components. Which of lives *b* and *c* has the higher welfare? There is a marginal loss for each g_1-component that has been exchanged for a g_2 component since g_2 is marginally worse than g_1. However, there is a bigger gain for each g_3-component that has been exchanged for a g_2-component. It is hard to deny that there is some *m* such that the smaller number of smaller losses is compensated for by the greater number of greater gains, and that in such cases, life *c* has higher welfare than life *b*. What about *a* and *c*? Since g_1 is

---

cannot be exceeded by a whole that consists of the objects of the same type. The value of the first element is 5 which is higher than the limit for the last element which is 4. Consequently, the first element is strongly superior to the last one. However, for each element in the sequence, its value is lower than the limit for the next object in the sequence. Consequently, no element in the sequence is strongly superior to the one that comes next.

[17] This was suggested by Wlodek Rabinowicz.

superior to g_2, it follows that life *a* has higher welfare than life *c*. Since *c* has higher welfare than *b*, and *b* has higher welfare than *a*, it follows by transitivity that *c* has higher welfare than *a*. Hence, we now have a contradiction: *a* has higher welfare than *c* and *c* has higher welfare than *a*. Consequently, a believer in Strong or Weak Superiority must deny one of the two compelling intuitions invoked in the reasoning above.[18]

6. PROOF OF C2 AND C3

Since Strong Superiority implies Weak Superiority, if we prove that C3 is true, then we have also proven that C2 is true.[19]

Proof: Assume that we have a finite descending sequence $g_1, \ldots, g_n$. Assume further that

(1) g_1 is weakly superior to g_n.

Since the sequence is finite, there must exist some element g_i which is the first element in the sequence such that g_1 is weakly superior to g_i. Let

(2) g_i be the first element in the sequence such that g_1 is weakly superior to g_i (i.e., there is no element g_j, $j < i$, such that g_1 is weakly superior to g_j).

It follows from the definition of Weak Superiority that

(3) there is a number *m* such that for any number *n*, *m* g_1-objects are more R than *n* g_i-objects.

Since g_i is the first element like this, it follows that g_1 is not weakly superior to g_{i-1}. Consequently,

(4) for some *k*, *m* g_1-objects are not more R than *k* g_{i-1}-objects.

Since we have assumed that the ordering R is complete, it follows that

(5) for some *k*, *k* g_{i-1}-objects are at least as R as *m* g_1-objects.

[18] Although the view that life *b* has higher welfare than life *a* is compelling in this context, the analogous view might not be convincing in other contexts, such as in population axiology. This step in the argument can be replaced, however, with weaker and more compelling conditions. For a discussion of this in connection with population axiology, see Arrhenius (2000), ch. 10.

[19] The proof here assumes completeness of the relation R. For a variant of this proof without completeness see Arrhenius and Rabinowicz (2003) and Arrhenius (2000), section 3.2 and 10.3-4.

By the transitivity of the relation R, (5) and (3) imply that

(6) for some k, k g_{i-1}-objects are more R than n g_i-objects.

Since, from (3), n can be any number, (6) means that g_{i-1} is weakly superior to g_i. Q.E.D.[20]

BIBLIOGRAPHY

Arrhenius, G. & Rabinowicz, W. "On Millian Discontinuities", mimeo, 2003.
Arrhenius, G. *Future Generations: A Challenge for Moral Theory*, F.D.-Diss., Uppsala: University Printers, 2000.
Brentano, F. "Loving and Hating" in Chisholm, R. M. (ed.) *The Origin of Our Knowledge of Right and Wrong*, New York: Routledge and Kegan Paul, 1969 [1907].
Broad, C. D. *Five Types of Ethical Theory*, London: Routledge and Kegan Paul, 1979 [1930].
Crisp, R. *Ideal Utilitarianism: Theory and Practice*, D.Phil. Diss., University of Oxford, 1988.
———. "Utilitarianism and the Life of Virtue"*, The Philosophical Quarterly*, vol. 42, no. 167, pp. 139-60, April 1992.
Edwards, R. B. *Pleasures and Pains: A Theory of Qualitative Hedonism*, Ithaca: Cornell University Press, 1979.
Glover, J. *Causing Death and Saving Lives*, London: Penguin Books, 1977.
Griffin, J. *Well-Being: Its Meaning, Measurement, and Moral Importance*, Oxford: Clarendon Press, 1986.
Hutcheson, F. *A System of Moral Philosophy,* Vol. 1, New York: Augustus M. Kelley, 1968 [1755].
Klint Jensen, K. *Om afvejning af værdier*, Ph.D. Diss., University of Copenhagen, 1996.
Lemos, N. M. "Higher Goods and the Myth of Tithonus", *The Journal of Philosophy*, vol. 90, Issue 9, September 1993.
McTaggart, J. M. E. *The Nature of Existence*, Cambridge: Cambridge University Press, 1927.
Mill, J. S. *Utilitarianism*, ed. Roger Crisp, Oxford: Oxford University Press, 1998.
Narveson, J. "Utilitarianism and New Generations", *Mind* 76, pp. 62-72, Jan 1967.

---

[20] I would like to thank Krister Bykvist, John Broome, Erik Carlson, John Skorupski, Howard Sobel, Wayne Sumner, Michael Zimmerman, and especially Wlodek Rabinowicz for very fruitful discussions and comments. Earlier versions of this paper were presented at the Philosophy Departments of the University of Bristol, Gothenburg University, and Lund University, and at the conferences *Choice and Value*, Corpus Christi College, Oxford University, *Utilitarianism, Human Rights and Globalization*, 7th International Conference of the International Society for Utilitarian Studies, Lisbon, Portugal, and at the Fourth Annual Bellingham Summer Philosophy Conference, Western Washington University. I would like to thank the participants at these occasions for their stimulating criticism, and especially Ben Bradley and Ellie Mason for their commentaries at the superb Bellingham Conference. The comments from two anonymous referees for *Philosophical Studies* were also useful. Thanks also to the Faculty of Philosophy, Oxford University, for hosting me during most of the time when this paper was written. Financial support from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) is gratefully acknowledged. Some of the ideas in this paper are also discussed in Arrhenius and Rabinowicz (2003).

Newman, J. H. *Certain Difficulties Felt by Anglicans in Catholic Teaching*, vol. I, London: Longmans and Green, 1885.
Parfit, D. *Reasons and Persons*, Oxford: Clarendon Press, 1984.
———. "Overpopulation and the Quality of Life", pp. 145-64 in P. Singer (ed.), *Applied Ethics*, Oxford: Oxford University Press, 1986.
Portmore, D. W. "Does the Total Principle Have Any Repugnant Implications?", *Ratio* vol. xii, March 1999.
Rabinowicz, W. "Ryberg's Doubts About Higher and Lower Pleasures – Put to Rest?", *Ethical Theory and Moral Practice* 6, pp. 231-7, 2003.
Rashdall, H. *The Theory of Good and Evil: A Treatise on Moral Philosophy*, vol. 1, 2nd ed., 1907.
Riley, J. "On Quantities and Qualities of Pleasure", *Utilitas* 5, pp. 291-300, 1993.
———. "Is Qualitative Hedonism Incoherent?", *Utilitas* 11, pp. 349-60, 1999.
Ross, W. D., *The Right and the Good*, Oxford: Oxford University Press, 1930.
Ryberg, J. "Higher and Lower Pleasures – Doubts on Justification", *Ethical Theory and Moral Practice* 5, pp. 415-29.
Sidgwick, H. *The Methods of Ethics*, 7th edn., London: Macmillan, 1967 [1907].
Skorupski, J. *Ethical Explorations*, Oxford: Oxford University Press, 1999.

CHAPTER 25

R. M. CHISHOLM

ORGANIC UNITIES

1. INTRODUCTION

G. E. Moore said of the first edition of *The Origin of Our Knowledge of Right and Wrong* [* a translation of Franz Brentano's *Vom Ursprung sittlicher Erkenntnis* that was published in 1902] that it is "a far better discussion of the most fundamental principles of ethics than any other works with which I am acquainted."[1] But he added that the book has this limitation: it "does not recognize, but even denies by implication, the principle I have called *the principle of organic unities*."[2] According to Moore's principle, "the value of a whole must not be assumed to be the same as the sum of the values of its parts."[3]

It is true that Brentano came to see the validity of this principle only after the publication, in 1889, of the first edition of *Vom Ursprung sittlicher Erkenntnis*. And it is also true that the principle of summation that he had affirmed in the lecture is inconsistent with the principle of organic unities.[4] Yet what Brentano there said about the "two unique cases of preferability" – namely, pleasure in the bad and displeasure in the bad – presupposes Moore's principle and, if I am not mistaken, enables us better to understand the concept of an organic unity.

In setting forth Brentano's views about organic unities, I will begin with those instances that he came to recognize after the publication of the first edition of *Vom Ursprung sittlicher Erkenntnis*.

---

[1] The quotation is from Moore's review of that work in the *International Journal of Ethics*, 14 (1903), pp. 115-123; see p. 115. This review is republished in Linda L. McAlister's *The Philosophy of Brentano* (London: Duckworth, 1976), pp. 176-181.
[2] *Principia Ethica* (Cambridge: Cambridge University Press, 1903), p. xi.
[3] Ibid., p. 28.
[4] He repudiated the principle of summation in his lectures on ethics that were given in the University of Vienna in 1894. Compare his *The Foundation and Construction of Ethics*, edited by F. Mayer-Hillebrand and translated by E. H. Schneewind (London: Routledge and Kegan Paul, 1973), pp. 194-197; this is a translation of *Grundlegung und Aufbau der Ethik* (Hamburg: Felix Meiner Verlag, 1978), pp. 211–214.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 305-318.
© 2005 *Springer. Printed in the Netherlands.*

2. THE IMPORTANCE OF ORDER

Writing on Leibniz, probably in the 1890s, Brentano observed that the value of a whole is not a function merely of the value of the parts of the whole. "Goodness also lies in the relations which are exhibited within the whole."[5] And in an unpublished manuscript, probably from the year 1908, Brentano wrote: "It is not merely the summation of the elements in an order that is to be considered as good; the order itself must be taken into consideration."[6]

Brentano cites three types of whole wherein the value may be said to be a function of the order of the elements. These include: (1) the *bonum variationis*; (2) the *bonum progressionis*; and (3) the value of retribution or requital. These all illustrate "the good that there is in order or arrangement."

(1) The principle of the *bonum variationis* might be formulated this way: other things being equal, it is better to combine two dissimilar goods than to combine two similar goods. Suppose, for example, that A is a beautiful painting, that B is a painting exactly like A and that C is a beautiful piece of music. The aesthetic contemplation of A may have the same value as that of B and also the same value as that of C. But the whole that is the aesthetic contemplation of A followed by that of C is intrinsically better than that whole that is the aesthetic contemplation of A followed by that of B.[7] Hence one could say that the value of a *bonum variationis* is greater than the sum of the values of its constituent parts.

(2) The principle of the "*bonum progressionis*" or the "*malum regressus*" might be put by saying: "If A is a situation in which a certain amount of value x is increased to a larger amount y, and if B is like A except that in B there is a decrease from the larger amount of value y to the smaller amount x, then A is preferable to B." Thus Brentano writes: "Let us think of a process which goes from good to bad or from a great good to a lesser good; then compare it with one which goes in the opposite direction. The latter shows itself as the one to be preferred. This holds even if the sum of the goods in the one process is equal to that in the other. And our preference in this case is one that we experience as being correct."[8] (In comparing the two processes, A and B, we must assume that each is the mirror image of the other. Hence the one should not include any pleasures of anticipation unless the other includes a corresponding pleasure of recollection.) The *bonum progressionis*, then, would be a good situation corresponding to A, in our formulation above, and the *malum regressus* would be a bad situation corresponding to B.

(3) Writing to Kraus in 1908, Brentano observed that one of the respects in which he had revised his original views had to do with "the law of retribution or

---

[5] From a manuscript entitled "Kritisches zu Leibniz's Optimismus," in Brentano's *Philosophie der Neuzeit*, ed. Klaus Hedwig (Hamburg: Felix Meiner, in press).

[6] "Nicht allein Summierung der Elemente der Ordnung ist als Gut zu betrachten, sondern auch die Ordnung selbst." The manuscript, listed in Brentano's *Nachlass* as M 18b, is entitled "Einteilung der Philosophie"; the passage cited is one of several "cosmological theses" set forth by Brentano.

[7] The example is adapted from Georg Katkov, *Untersuchungen zur Werttheorie und Theodizee* (Brünn: Verlag Rudolf M. Rohrer, 1967), p. 56.

[8] *Foundation*, pp. 196-197; *Grundlegung*, p. 214.

requital [*Vergeltungssatz*], concerning which Leibniz makes some valuable observations in his *Theodicy*."[9] Brentano had observed, in the essay "Loving and Hating," which was dictated in 1907, that "if at the Last Judgment a greater amount of bliss were given to a person who actually deserved it less, then he would have a greater amount of good than he otherwise would have, but the good in the universe considered as a whole would be less."[10]

In an unpublished fragment, entitled "On the good that there is in order or arrangement," Brentano says that the evil that is involved in retribution may yet make a bad situation less bad than it would have been without the retribution. Wickedness accompanied by sorrow is better than the same wickedness accompanied by pleasure; this fact, Brentano suggested, may justify the sorrow that is involved in repentance and the pain that is involved in vindictive punishment. If A is a wicked deed and if B is the suffering involved in the sinner's remorse or in his retribution, then the two evils, A and B, may be preferable to A without B. Brentano writes:

> (1) Given the same amount of pleasure, displeasure, moral goodness, and moral evil, it would seem that the world is better if the pleasure is associated with the good and the displeasure associated with the evil than in the converse case; and it is best when these are in proportion.
> (2) Indeed, moral evil with suffering would seem to be better than the same moral evil with pleasure.
> (3) As a result of this, the principle of vindictive punishment acquires a certain significance.
> (4) And the sorrow of repentance is justified. (Atonement.)

Brentano adds that the failure to consider such points as these constitutes a significant gap in the original edition of *The Origin of Our Knowledge of Right and Wrong*.[11]

Yet what Brentano had to say in his original lecture of 1889 about *pleasure in the bad* and *displeasure in the bad* provides us with perhaps the clearest examples of what Moore had called organic unities – wholes that are not "mere sums."

---

[9] F. Brentano, *The Origin of Our Knowledge of Right and Wrong*, translated by R. M. Chisholm and E. H. Schneewind (London: Routledge and Kegan Paul, 1969), p. 115; this is a translation of *Vom Ursprung sittlicher Erkenntnis*, edited by O. Kraus, third edition (Hamburg: Felix Meiner Verlag, 1934), p. 114. Compare Leibniz's *Theodicy*, 1, para. 73.

[10] *Origin*, p. 149; *Ursprung*, p. 155.

[11] "(1) Bei gleichem Mass von Lust, Unlust, sittlicher Gute und Schlechtigkeit erscheint die Welt besser, wo die Lust den Guten und die Unlust den Schlechten zukommt als umgekehrt, am besten die, wo in Proportion. (2) Ja, es scheint gleiche Schlechtigkeit mit Leid besser als gleiche Schlechtigkeit mit Lust. (3) Hieraus bekommt das vindikative Strafprinzip eine gewisse Bedeutung. (4) Auch der Reueschmerz erhält Berechtigung. (Busserleben)..." The fragment, listed in Brentano's *Nachlass* as Ethik 1, is entitled "Vom Guten, das in der Zuordnung liegt." Brentano adds that these considerations are relevant to what may be the oldest proof of the immortality of the soul – that of Socrates, which is based upon the justice of God. "Is there hell, then? I think and hope: No!"

3. MERE SUMS

What is it for a whole to be a "mere sum"?

I will follow Moore's terminology and say that the "wholes" in question are states of affairs that have other states of affairs as "parts." The relevant sense of this new, value use of "part" is this:

> P is a part of Q = Df Q is necessarily such that (a) if it obtains then P obtains and (b) whoever conceives it conceives P.

(We could replace the definiens "P is a part of Q" by "Q entails P.") We are here using "part" to mean the same as "proper part." Using "part" in this way, we cannot say of any whole that it is a part of itself.

We must distinguish the two questions "What is a mere sum?" and "Which wholes are mere sums?" The former is a *logical* question about the nature of value, the latter is a *material* question. Since our present concern is to answer the logical question, it will be useful to begin by oversimplifying some of the issues that are involved in answering the material question.

Let us, for the present, adopt the fiction that hedonism is the correct theory of value: states of pleasure are the only things that are intrinsically good and states of displeasure or pain are the only things that are intrinsically bad. Let us also pretend that, of two pleasures, the one that is the more intense is of greater value and that, of two displeasures, the one that is the more intense is of lesser value. Let us pretend, further, that we can measure pleasures and displeasures quantitatively: we can say of a pleasure or displeasure that it has a value of *n* if and only if it has an intensity of *n*. We may now represent the positive value of a state of affairs with a positive number and the negative value of a state of affairs with a negative number.

Now we may consider a number of wholes that would be "mere sums" if our hedonistic assumptions were true. The value of each such whole would be "merely a sum of the values of its parts." There are six types of such whole: (1) those implying pleasure but no displeasure; (2) those implying displeasure but no pleasure; (3) those implying neither pleasure nor displeasure; (4) those implying both pleasure and displeasure but more of the former than the latter; (5) those implying both pleasure and displeasure but more of the latter than the former; and (6) those implying both pleasure and displeasure and equal amounts of both.

These six "sums" may be represented in the following diagram. The value of the good parts, if any, is represented in the first line; that of the bad parts, if any, is represented in the second line; and the value of the whole, which is a mere sum of the values of the good and the bad parts, is represented in the third line.

	(1)	(2)	(3)	(4)	(5)	(6)
(G)	2	0	0	2	1	1
(B)	0	-2	0	-1	-2	-1
(W)	2	-2	0	1	-1	0

Moore had said that a "mixed whole" is a whole having both good and bad parts. (We may assume that every whole, whether good or bad, has neutral parts.) He also said that a "pure good" is a good whole having no bad parts and that a "pure evil" is a bad whole having no good parts. Using Moore's terminology, we may say: (1) is a pure good; (2) is a pure evil; (3) is a pure neutral; (4) is a mixed good; (5) is a mixed evil; and (6) is a mixed neutral. (Note that a mixed neutral is a neutral having both good and bad parts. If all wholes were mere sums, then every *unmixed* neutral would be a pure neutral – just as every unmixed good is a pure good and every unmixed evil is a pure evil. But some organic unities are unmixed neutrals that are not pure neutrals: that is to say, either they have a bad part and no good part or they have a good part and no bad part.) Consider now the mixed wholes that we have distinguished.

In the mixed good (4), the badness of the part that is bad is *outweighed* by the part that is good. In the mixed evil (5), the goodness of the part that is good is *outweighed* by the part that is bad. And in the mixed neutral (6), the goodness of the good part and the badness of the bad part *counterbalance* each other.

Such "mere sums of value" may now be contrasted with *organic unities*. To define the latter concept, we introduce a definition of a "mixed whole":

> W is a mixed whole = Df Either (a) W has both good and bad parts or (b) W is neutral and has either a good or a bad part.

We will say that a whole is "unmixed" if it is not mixed.

> W is an organic unity = Df Either (a) W is mixed and has no unmixed parts such that it falls in value between them or (b) W is unmixed and has a good part that is better than it or a bad part that is worse than it.

One thing may be said to "fall in value between" two others provided it is better than one of them and worse than the other.

I will distinguish ten types of organic unity.

4. DEFEAT

We begin with what may be called the "defeat" of intrinsic value. This concept may be illustrated by the "unique cases of preferability" that Brentano has discussed: pleasure in the bad; displeasure in the bad; pleasure in the good; and displeasure in the good.

According to Brentano's theory of value, (1) pleasure is intrinsically good, (2) displeasure is intrinsically bad, (3) consciousness is intrinsically good, and (4) consciousness has no contrary that is intrinsically bad. Hence, *someone experiencing pleasure* will be intrinsically good, *someone experiencing displeasure* will be

intrinsically bad, and *someone being conscious,* which is part of each, will be intrinsically good.

Let us think of "joy" and "sorrow" as designating species of pleasure and displeasure, respectively. And let us understand "joy" and "sorrow" intentionally; that is to say, we will interpret them as referring to emotional attitudes that may have propositional objects that are false. Thus "Jones takes joy in Smith being happy" will not imply that Smith is happy, and "Smith takes sorrow in Jones being unhappy" will not imply that Jones is unhappy.

Meinong used the following terms to express Brentano's distinctions: *Mitfreude*, or joy in the other person's joy; *Mitleid*, or sorrow in the other person's sorrow; *Schadenfreude*, or joy in the other person's sorrow; and envy (*Neid*), or sorrow in the other person's joy.[12]

Suppose now we say that, though joy is good and sorrow is bad, nevertheless *Schadenfreude,* or that joy that is joy in another person's sorrow, is neutral, and that envy, or that sorrow that is sorrow in another person's joy, is also neutral. Joy in another person's sorrow, then, would have a good part that is better than it, but it would not have a bad part that is worse than it. And sorrow in another person's joy would be a neutral whole having a good part that is better than it and no bad part that is worse than it.

Before considering these examples in further detail, we should remind ourselves that, given our definition of what it is for something to be a "part" of a state of affairs, we cannot say that the *correctness* or *incorrectness* of an emotion is a "part" of that emotion. If P is a part of a state of affairs Q, then Q is necessarily such that whoever considers it considers P. One cannot consider *Jones taking pleasure in the good* without considering *Jones taking pleasure.* But one can consider *Jones taking pleasure in the good* without considering *Jones's pleasure being correct.* This accords with Brentano's conception. The *correctness* of an emotion, according to him, is not one of those things that contribute to the value of an emotion. For to say that an emotion *has* intrinsic value is only to say that the emotion *is* correct.[13]

Let us now consider *Schadenfreude* in more detail.

What is bad about pleasure in the bad? The answer is suggested by what Schopenhauer says about "malignant joy": "In a certain sense the opposite of envy is the habit of gloating over the misfortunes of others. At any rate, whereas the former is human, the latter is diabolical. There is no sign more infallible of an entirely bad heart, and of profound moral worthlessness than open and candid enjoyment in seeing other people suffer."[14] In this case, we would have a mixed evil that has no part that is worse than itself.

---

[12] A. Meinong, *Gesamt Ausgabe*, vol. III (Graz: Akademische Druck- und Verlagsanstalt, 1968), pp. 714-718.
[13] For a discussion of this general question, compare Lynn Pasquerella. "Brentano and Organic Unities," *Proceedings of the Tenth International Wittgenstein Symposium* (Vienna: Hölder-Pichler-Tempsky, 1985).
[14] A. Schopenhauer, *The Basis of Morality* (London: Swan Sonnenschein, 1903), pp. 156-157.

Pleasure in the bad has a proper part that is good: someone taking pleasure in something. Does it have a proper part that is bad? If we view pleasure in the bad as Brentano does, then we must say that it logically implies another state of affairs that is bad: someone experiencing an emotion that is incorrect. But it does not follow that experiencing pleasure in the bad has experiencing an incorrect emotion as a *part.* We have said that a state of affairs P is a *part* of a state of affairs Q if and only if the following conditions hold: Q is necessarily such that (a) if it obtains then P obtains and (b) whoever conceives it conceives P. One can contemplate experiencing pleasure in the bad without thereby contemplating experiencing an emotion that is incorrect.

If pleasure in the bad illustrates the defeat of goodness, then displeasure in the bad may illustrate the defeat of badness. Aristotle refers to "pain at unmerited good fortune."[15] Righteous indignation provides another example. Saint Thomas defines the righteously indignant man as one who is "saddened at the prosperity of the wicked."[16] If we are to have an example of an organic unity, we must modify Saint Thomas's definition and say that the righteously indignant man is one who is saddened at what he *takes* to be the prosperity of the wicked. That he takes the wicked to be prospering, we might say, is a state of affairs that is intrinsically good; we could say this on the ground that an act of consciousness is intrinsically good. Or, if we prefer not to go this far, we may say that the state of consciousness is intrinsically neutral. That the person is *saddened* is a state of affairs that is intrinsically bad. The wider state of affairs that is the person's being saddened at what he takes to be the prosperity of the wicked, is, according to what Saint Thomas suggests, better than the narrower state of affairs that is the person's being conscious (whether or not the narrower state of affairs is good or neutral); and it is better just because of the presence of the part that is bad. The wider state of affairs, then, has a part that is good or neutral and a part that is bad. It is itself a good state of affairs, but unlike what we found in the case of mere balancing off, it has no part that is *better* than it.[17]

If evil is *balanced off* in a larger whole, then we may regret or resent the presence of the evil in that whole. But if these examples of *defeat* of evil are acceptable, then we should be thankful for the badness of the part that is bad. For in

---

[15] "Most directly opposed to pity is the feeling called Indignation. Pain at unmerited good fortune is, in one sense, opposite to pain at unmerited bad fortune, and is due to the same moral qualities. Both feelings are associated with good moral character; it is our duty both to feel sympathy and pity for unmerited distress, and to feel indignation at unmerited prosperity; for whatever is undeserved is unjust, and that is why we ascribe indignation even to the gods." *Rhetoric,* bk. II, chap. 9, 1386b.

[16] Saint Thomas, *Commentary on the Nichomachean Ethics*, para. 356. See Aristotle's *Ethics* II.7.1108b.

[17] The example would illustrate the defeat of badness even without the assumption that consciousness is, as such, intrinsically good. Thus the wider state of affairs, which is the person being saddened by the prosperity of the wicked, could be said to be good; it has a bad part, which is the person being sad; but it has no part that is better than it.

each case the badness of the part that is bad makes the whole *better* than what we would have if the bad part had been replaced by its neutral negation.

Defeat may be illustrated in still other ways.

The unpleasant experience of fear, we may suppose, is a state of affairs that is intrinsically bad. But such experience is necessarily involved in the exercise of courage. And the exercise of courage, we may further suppose, is a virtuous activity that is intrinsically good. We need not pause to consider what else it is that goes with fear to make up courage. For the point of the present example is that the larger whole – the exercise of courage – is *better* intrinsically because of the badness of the part that is bad. It has been held, indeed, that all virtue requires the presence of evil. Thus Lactantius, in dealing with the problem of evil, said that, if evil were removed, then "no trace of virtue would remain, since virtue consists of bearing with and overcoming the pains of evil."[18]

Consider now an example from aesthetics. A certain combination of paints may be ugly. This combination may be entailed by a larger whole that is not ugly or that is even beautiful. And the larger whole may be preferable aesthetically just because of the ugliness of the part that is ugly.

This example suggests that the concept of defeat provides the solution to a problem in aesthetics that Brentano had discussed at length in 1892. As we have seen, Brentano holds that the value of any act of thought – of any presentation – is in part a function of the value of the *object* of that presentation. To think about what is good is better than to think about what is bad. But if we leave the matter there we are confronted with what would seem to be a philosophical paradox. Brentano discusses this apparent paradox, or "riddle," in a lecture entitled "The Bad as Object of Poetic Representation," given in Vienna before the "Gesellschaft der Literaturfreunde" in 1892, three years after the publication of the first edition of the *Ursprung*.[19]

The riddle might be put this way. "Aesthetic contemplation is unquestionably intrinsically valuable. But the objects that are depicted in the various arts need not be intrinsically good. We have only to consider tragedy and comedy in order to realize that, at their best, the arts depict to us things that are intrinsically bad. But how can this be if the value of a presentation is a function of the value of its object?"

The solution, Brentano says, is that "the special value of a representation is not *solely and alone* a function of the goodness of that which is represented."[20] Although he is not entirely clear about what the other factors are that lend value to the object of poetic representation, it is evident that, according to him, the

---

[18] Lactantius also said that, if evil were removed, "wisdom would also be removed" (since "unless we first know evil we shall not be able to know good"), and "there is more good and satisfaction in wisdom than there is painfulness in evil." Quoted by Pierre Bayle in Note E of his "Paulicians"; in Richard Popkin, ed., Pierre Bayle, *Historical and Critical Dictionary* (Indianapolis: Bobbs Merrill, 1965), p. 169.

[19] The lecture, called "Das Schlechte als Gegenstand dichterischer Darstellung," was published separately in 1892 and is reprinted in the *Grundzüge der Aesthetik*, pp. 170-195.

[20] Ibid., p. 177.

presentation of a tragedy or a comedy involves a whole in which the evil of certain proper parts has been defeated.

5. AN OBJECTION TO THE CONCEPT OF DEFEAT

One may object: "The examples you cite of defeated goods and evils are all states of affairs that are indeterminate. Thus you cite *Jones being pleased* as a good which is defeated by the larger whole *Jones being pleased that Smith is displeased.* But *Jones being pleased* is indeterminate with respect to the intentional object of Jones's pleasure whereas the larger whole is not. Evidently it is only by restricting yourself to such indeterminate states of affairs that you can find any examples of defeated goods or evils. But the predication of intrinsic goodness and intrinsic evil should be restricted to those states of affairs that are completely determinate. And you will find that, when it is so restricted, then there will be no cases of the defeat of intrinsic value."

What does it mean to say of a state of affairs that it is "indeterminate"? One may mean to be saying of the state of affairs that it is an "incomplete object" – that there is some property that is such that the state of affairs neither has that property nor fails to have that property.[21] But in this sense of "indeterminate," the state of affairs that is *Jones being pleased* is *not* indeterminate; like everything else, it is such that, for every property, either it has that property or it fails to have the property.

In saying that *Jones being pleased* is "indeterminate," one may mean, however, something less implausible. One may mean that the state of affairs itself has *no* intrinsic value and that intrinsic value can be assigned only to certain more encompassing states of affairs of which the given state of affairs is only a proper part.[22] But how are we to decide which states of affairs are sufficiently encompassing to merit predication of intrinsic value? Every state of affairs, after all, is a part of a more encompassing state of affairs – *except* for those all-encompassing states of affairs that are sometimes called "worlds." But the assumption that intrinsic value can be attributed only to worlds is inconsistent with our presupposition according to which there are some things we know to be good and some things we know to be bad.

6. THE LOGICAL PROPERTIES OF DEFEAT

We may now distinguish more precisely the several senses of defeat.

---

[21] This is the way Meinong characterizes incomplete objects. See *Über Möglichkeit und Wahrscheinlichkeit* in *Gesamtausgabe*, vol. 6 (Graz: Akademische Druck- und Verlagsanstalt, 1972), pp. 178 ff.

[22] This is the view of Michael Zimmerman, who suggests that certain states of affairs are "evaluatively incomplete" and therefore neither good, bad, nor neutral. But he does not propose any way of deciding whether a state of affairs is evaluatively incomplete. See Michael J. Zimmerman, "Evaluatively Incomplete States of Affairs," *Philosophical Studies,* 43 (1983), pp. 211-224.

The definition of the defeat of goodness should be adequate to the three types of pleasure in the bad that we have distinguished: (1) the case where pleasure in the bad is a bad whole with a good part; (2) the case where pleasure in the bad is a neutral whole, having a good part and no bad part; and (3) the case where pleasure in the bad is a good whole, having a part that is better than it and no part that is bad.

We first consider the defeat of goodness. We will distinguish the *partial defeat* of goodness, the *total defeat* of goodness, and the *transvaluation* of goodness.

We begin with the concept expressed by saying, "Some of the goodness of a good state of affairs is defeated by a larger whole in which it appears":

> Some of the goodness of G is defeated by W = Df G is a good part of W and better than W; and if W has a bad part that is worse than W, then that bad part is part of G.

There are three ways in which goodness may be defeated. The goodness of G is *partially defeated* by the wider whole W, provided that some of the goodness is defeated and that the wider whole W is good. The goodness of G is *totally* defeated provided that some of the goodness is defeated and the wider whole is *not* good. And the goodness of G is *transvalued* by W provided some of the goodness of W is defeated by W and W is bad.

The merely *partial* defeat of goodness is illustrated by pleasure in the bad – if such pleasure is good, has a part (pleasure) that is no better than it, and has no part that is bad.

The *total* defeat of goodness is also illustrated by pleasure in the bad – if such pleasure is neutral, has a good part (pleasure), and has no proper part that is bad.

The *transvaluation* of goodness is illustrated by pleasure in the bad – if such pleasure is bad, has a good proper part (pleasure), and has no proper part that is bad.

The defeat of evil is analogous. We begin with this abbreviation:

> Some of the badness of B is defeated by W = Df B is a bad part of W and worse than W; and if W has a good part that is better than W, then that good part is part of B.

The defeat of evil, like the defeat of goodness, may take three forms. The evil in the larger whole is only *partially defeated* if the larger whole is itself bad. The evil is *totally defeated* if the whole is *not* bad. And the evil is *transvalued* if the whole is good.

The merely *partial* defeat of evil is illustrated by displeasure in the bad – if such displeasure is bad, and has a part (displeasure) that is worse than it. Another example is the punishment of a wicked deed – if it is better than any of its bad proper parts.

The *total* defeat of evil is illustrated by displeasure in the bad – if such displeasure is neutral and has a part (displeasure) that is worse than it.

The *transvaluation* of evil is illustrated by displeasure in the bad – if such displeasure is good and has a proper part that is bad.

When evil is balanced off by a larger whole, then the larger whole has a good part that is not a part of the bad part and that is better than the whole itself. But when evil is defeated, the larger whole will not have any good part that is not a part of the bad part. Hence the whole will not have outside of the bad part any part that is better than the whole itself.

And so when the badness of a state of affairs p is defeated by a larger whole, it will not be true that the badness of p is balanced off by that larger whole; we will not find elsewhere in the whole a part that is good and better than the whole. When badness is balanced off, and not defeated, by a whole that is good, then one may yet regret or resent its presence there. But if badness is ever defeated by a whole that is good, then, as I have suggested, we may well be thankful for the very presence of the part that is bad.

7. THE POSITIVE AND NEGATIVE DEFEAT OF THE NEUTRAL

We have spoken only of the defeat of goodness and the defeat of evil. But we may also distinguish the defeat of neutrality.

We have the defeat of the neutral when we have a whole W of the following sort: W is not neutral but all its proper parts are neutral. The defeat of the neutral is *positive* when W is good and *negative* when W is evil. We cite an example of each.

The *negative* defeat of the neutral is illustrated in this situation: a person is aware of what he takes to be his evil deed and is not ashamed. The whole is bad but has no proper parts that are bad.

Aristotle says that "shamelessness – not to be ashamed of doing bad actions – is bad."[23] Suppose I believe I have performed an evil deed and I contemplate what I take to be this deed. The contemplation of this deed – of this supposed deed – will not itself be bad; we may say that it is neutral, or even that it is good. Let us take the former course. Consider now that neutral state of affairs that is my not being ashamed of that which I contemplate. If we combine these two neutrals we arrive at that bad state that is failure to be ashamed at the contemplation of one's misdeed.

The *positive* defeat of the neutral is illustrated in this situation: a person is aware of what he or she takes to be someone wronging him and is not resentful. The whole is good but has no proper parts that are good.

[23] *Nicomachean Ethics,* IV, chap. 9, 1128b.

8. ENHANCEMENT AND AGGRAVATION

I now turn to organic unities of a somewhat different sort. These illustrate what may be called "the enhancement of goodness" and "the aggravation of evil."[24]

We might characterize *enhancement* somewhat loosely by saying: "A good may be said to be enhanced when, without being combined with any second good, it is part of a whole that is better than it." And we might characterize *aggravation* somewhat loosely by saying: "An evil may be said to be aggravated when, without being combined with any second evil, it is part of a whole that is worse than it."

The following is a more precise definition of enhancement:

> W enhances the goodness of G = Df G is part of W; G is good but not as good as W; and every proper part of W that is good has a part in common with G.

(It should be recalled that, if P is a "proper part" of Q, then, although P is a part of Q, Q is not a part of P.) The definition of aggravation is analogous to that of enhancement:

> W aggravates the value of B = Df B is part of W; B is bad but not as bad as W; and every proper part of W that is bad has a part in common with B.

Schopenhauer, in attempting to reconcile his pessimism with the presence of goodness in the world, notes that goodness may serve to aggravate evil.

Enhancement would be illustrated by pleasure in the good – if (1) it has pleasure as its only good proper part and (2) it is itself better than its good proper part. The *bonum progressionis* is also an instance of enhancement.

There is an instructive discussion of enhancement by Saint Thomas.[25] He considers the following argument that is designed to show that the blessed in heaven will not see the sufferings of the damned. "The Philosopher says that 'the most perfect operation of the sense of sight is when the sense is most disposed with reference to the most beautiful of the objects which fall under the sight' (*Ethics,* x 4). Therefore, on the other hand, any deformity in the visible object redounds to the imperfection of the sight. But there will be no imperfection in the blessed. Therefore they will not see the sufferings of the damned wherein there is extreme deformity."

Saint Thomas then makes this comment: "Nothing should be denied the blessed that belongs to the perfection of their beatitude. Now everything is known the more for being compared with its contrary, because when contraries are placed beside one

---

[24] Fred Feldman first called my attention to the fact that enhancement and aggravation, as here understood, are to be distinguished from defeat.

[25] See *Summa Theologica*, suppl., Q. 94 ("Of the Relations of the Saints toward the Damned"), art. 1 ("Whether the Blessed in Heaven Will See the Sufferings of the Damned"). The translation used here is from vol. III of Saint Thomas Aquinas, *Summa Theologica* (New York: Bensiger Brothers, 1948).

another they become more conspicuous. Wherefore in order that the happiness of the saints may be more delightful to them [*ut beatitudo sanctorum eis magis complaceat*] and that they may render more copious thanks to God for it, they are allowed to see perfectly the sufferings of the damned."[26]

And now he replies to the objection: "Although the beauty of the thing seen conduces to the perfection of vision, there may be deformity of the seen without imperfection of vision: because the images of things whereby the soul knows contraries are not themselves contrary. Wherefore also God who has most perfect knowledge sees all things, beautiful and deformed."

The third article of Question 94 is entitled: "Whether the Blessed Rejoice in the Punishment of the Wicked?" Saint Thomas answers the question affirmatively and then considers the objection: "…rejoicing in another's evil pertains to hatred. But there will be no hatred in the blessed."

Saint Thomas refers to Scripture: "The just shall rejoice when he shall see the revenge" (*Psalms* lvii, 11). He then observes: "A thing may be a matter of rejoicing in two ways. First, directly [per se], when one rejoices in a thing as such; and thus the saints will not rejoice in the punishment of the wicked. Secondly, indirectly [per accidens], by reason namely of something annexed to it; and in this way the saints will rejoice in the punishment of the wicked, by considering therein the order of Divine Justice and their own deliverance, which will fill them with joy. And thus the Divine Justice and their own deliverance will be the direct cause of the joy of the blessed [*per se causa gaudii beatorum*]; while the punishment of the damned will cause it indirectly [per accidens]."

And so he replies to the objection: "To rejoice in another's evil as such belongs to hatred, but not to rejoice in another's evil by reason of something annexed to it. Thus a person sometimes rejoices in his own evil as when we rejoice in our own afflictions as helping us to merit life: 'My brethren, count it all joy when you shall fall into divers temptations' (*James* i, 2)."

9. THE TYPES OF ORGANIC UNITY: A SUMMARY

For convenience, I now list definitions of the ten types of organic unity that have been distinguished. I first cite three preliminary definitions:

D1 P is a part of Q = Df Q is necessarily such that (a) if it obtains then P obtains and (b) whoever conceives it conceives P.

D2 Some of the goodness of G is defeated by W = Df G is a good part of W and better than W; and if W has a bad part that is worse than W, then that bad part is part of G.

[26] "… says he, gently as a lamb [sagt er sanft wie ein Lamm]"; Friedrich Nietzsche, *The Genealogy of Morals* (New York: Boris and Liveright, 1948), Essay I, 14.

D3 Some of the badness of B is defeated by W = Df B is a bad part of W and worse than W; and if W has a good part that is better than W, then that good part is part of B.

The ten types of organic unity are these:

OU1 The goodness of G is partially defeated by W = Df Some of the goodness of G is defeated by W, and W is good.

OU2 The goodness of G is totally defeated by W = Df Some of the goodness of G is defeated by W, and W is not good.

OU3 The goodness of G is transvalued by W = Df Some of the goodness of G is defeated by W, and W is bad.

OU4 The badness of B is partially defeated by W = Df Some of the badness of B is defeated by W, and W is bad.

OU5 The badness of B is totally defeated by W = Df Some of the badness of B is defeated by W, and W is not bad.

OU6 The badness of B is transvalued by W = Df Some of the badness of B is defeated by W, and W is good.

OU7 W positively defeats its neutrality = Df W is good, and every proper part of W is neutral.

OU8 W negatively defeats its neutrality = Df W is bad, and every proper part of W is neutral.

OU9 W enhances the goodness of G = Df G is part of W; G is good but not as good as W; and every proper part of W that is good has a part in common with G.

OU10 W aggravates the value of B = Df B is part of W; B is bad but not as bad as W; and every proper part of W that is bad has a part in common with B.

CHAPTER 26

N. M. LEMOS

CHISHOLM'S DEFINITION OF ORGANIC UNITY

I wish to consider briefly Chisholm's attempt to define an organic unity. One merit of his approach is that it does not make use of the notion of a "sum" of values. In *Brentano and Intrinsic Value,* he offers the following definitions:[1]

(D1) *P* is a part of *Q* = *Df.* *Q* is necessarily such that (a) if it obtains then *P* obtains, and (b) whoever conceives *Q* conceives *P*.

(D2) Some of the goodness of *G* is defeated by *W* = *Df.* *G* is a good part of *W* and better than *W*; and if *W* has a bad part that is worse than *W*, then that bad part is part of *G*.

(D3) Some of the badness of *B* is defeated by *W* = *Df.* *B* is a bad part of *W* and worse than *W*; and if W has a good part that is better than *W*, then that good part is part of *B*.

Given the definition of a part in (D1), every state of affairs is a part of itself. Thus, in (D2) and (D3) we should take the definition of "part" to refer to "proper parts." *P* is a proper part of *Q* if and only if *P* is a part of *Q* and *Q* is not a part of *P*.

The defeat of goodness can be illustrated by pleasure in the bad. Consider the state of affairs, Smith's being pleased that Jones is suffering. Let us assume that this state of affairs is bad. This state of affairs has no bad parts, for Jones's suffering is *not* a part of it. (Smith's being pleased that Jones is suffering does not imply that Jones is suffering.) Still, it has a good part which is better than it, namely, Smith's being pleased. The defeat of badness can be illustrated by displeasure in the bad. Consider the state of affairs, Smith's being displeased that Jones is suffering. Let us assume that this state of affairs is good or neutral. It has a bad part which is worse than it, namely, Smith's being displeased. It has no good part.

---

[1] Roderick Chisholm, *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press, 1986), p. 88 [* pp. 317-18 of this volume].

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 319-323.
© 2005 *Springer. Printed in the Netherlands.*

Given these definitions Chisholm distinguishes the following ten types of organic unities:[2]

OU1 The goodness of *G* is partially defeated by *W* = *Df.* Some of the goodness of *G* is defeated by *W*, and *W* is good.

OU2 The goodness of *G* is totally defeated by *W* = *Df.* Some of the goodness of *G* is defeated by *W*, and *W* is not good.

OU3 The goodness of *G* is transvalued by *W* = *Df.* Some of the goodness of *G* is defeated by *W*, and *W* is bad.

OU4 The badness of *B* is partially defeated by *W* = *Df.* Some of the badness of *B* is defeated by *W*, and *W* is bad.

OU5 The badness of *B* is totally defeated by *W* = *Df.* Some of the badness of *B* is defeated by *W*, and *W* is not bad.

OU6 The badness of *B* is transvalued by *W* = *Df.* Some of the badness of *B* is defeated by *W*, and *W* is good.

OU7 *W* positively defeats its neutrality = *Df.* *W* is good, and every proper part of *W* is neutral.

OU8 *W* negatively defeats its neutrality = *Df.* *W* is bad, and every proper part of *W* is neutral.

OU9 *W* enhances the goodness of *G* = *Df.* *G* is part of *W*; G is good but not as good as *W*; and every proper part of *W* that is good has a part in common with *G*.

OU10 *W* aggravates the value of *B* = *Df.* *B* is part of *W*; *B* is bad but not as bad as *W;* and every proper part of *W* that is bad has a part in common with *B*.

Finally, Chisholm offers the following definitions;[3]

(D4) *W* is a mixed whole = *Df.* Either (a) *W* has both good and bad parts or (b) *W* is neutral and has either a good or a bad part.

(D5) *W* is an organic unity = *Df.* Either (a) *W* is mixed and has no unmixed parts such that it falls in value between them or (b) *W* is unmixed and has a good part that is better than it or a bad part that is worse than it.

What would be examples of a mixed whole? The following is a mixed whole: *A*'s experiencing pleasure and *B*'s experiencing pain. This is a mixed whole because it

[2] Ibid., pp. 88-9 [* p. 318 of this volume].
[3] Ibid., p. 75 [* p. 309 of this volume].

has both good and bad parts. A second example of a mixed whole would be any neutral state of affairs in which the goodness of *A*'s experiencing pleasure was precisely balanced by the badness of *B*'s experiencing pain. (If we were to pretend that quantitative hedonism is true and that we could measure quantitatively the intensity of pleasures and pains, an example of this latter state of affairs would be *A*'s having six units of pleasure and *B*'s having six units of pain.) Such a neutral state of affairs would have a good part and a bad part.

If we assume that the state of affairs *Smith's being pleased that Jones is suffering* is intrinsically bad, then it is also an unmixed whole, since it has no bad parts. According to (D5), it is also an organic unity because it is an unmixed whole that has a good part, namely, Smith's being pleased, which is better than it is. Similarly, if we assume that the state of affairs *Smith's being displeased that Jones is suffering* is intrinsically good, then it is also an unmixed whole, since it has no good parts. According to (D2), it is an organic unity because it is an unmixed whole and has a bad part that is worse than it is. Chisholm's definition of an organic unity is at least adequate to these examples. Unfortunately, Chisholm's definition of an organic unity is not adequate to certain forms of organic unity that he has distinguished. It is not adequate to (OU7), (OU8), (OU9), or (OU10).

What is an example of the "negative defeat of the neutral"? Chisholm suggests one's being aware of what one takes to be one's evil deed and not being ashamed.[4] This whole, he suggests, is intrinsically bad but has no parts that are bad. The "positive defeat of the neutral" is illustrated by one's being aware of what he takes to be someone wronging him and not being resentful. Chisholm says that the concept of "enhancement" may be illustrated by pleasure in the good if we suppose that *A*'s being pleased that someone is happy (*W*) is better than its simpler part, *A*'s being pleased (*G*). Similarly, we might say that an example of "aggravation" is *A*'s being displeased that someone is pleased if we assume that this whole is worse than its simpler part, *A*'s being displeased.

To see that those organic unities that "positively defeat their neutrality" (OU7) are not captured by Chisholm's definition of an organic unity, we need only note that every whole of this sort is unmixed. They are unmixed because they have only neutral parts. The fact that these are unmixed wholes having only neutral parts is the reason they cannot be captured by (D5). According to (D5), an unmixed whole is an organic unity only if it has a good part that is better than it or a bad part that is worse than it. Obviously if a whole has only neutral parts, it can't have a good part or a bad part. The same problem pertains to those organic unities that "negatively defeat their neutrality" (OU8). These wholes are also unmixed, since they have only neutral parts. But again, (D5) tells us that an unmixed whole is an organic unity only if it has a good part better than it or a bad part worse than it. So Chisholm's definition isn't adequate to these wholes either.

---

[4] Chisholm cites Aristotle's remark in the *Nicomachean Ethics:* "shamelessness – not to be ashamed of doing bad actions – is bad." Book IV, Chapter 9, 1128b. Shamelessness is an example of indifference toward the bad.

Definition (D5) is also not adequate to those wholes that are examples of enhancement. Suppose that the whole that is *A*'s being pleased that someone is happy, *W*, is a case of enhancement, that *W* enhances the goodness of *A*'s being pleased, *G*. In this case, *W* is an instance of enhancement, since it is better than its only good part, namely, *G*. Now *W* is an unmixed whole because it is a good whole with no bad parts. But according to (D5), *W* is an organic unity only if it has a good part that is better than it or a bad part that is worse than it. Yet *W* does not have a good part that is better than it or a bad part that is worse than it. So Chisholm's definition of an organic unity is not adequate to this case of enhancement.

Furthermore, (D5) is not adequate to certain examples of aggravation. Let's assume that the whole *A's being displeased that someone is happy*, *W*, is an instance of aggravation, that *W* aggravates the badness of *A*'s being displeased, *B*. Let's assume that *W* is worse than its only bad part, *B*. *W* is an unmixed whole, since it is bad with no good parts. But again, (D5) claims that an unmixed whole is an organic unity only if it has a good part that is better than it or a bad part that is worse than it. Yet *W* has no good part that is better than it and no bad part that is worse than it.

If Chisholm's definition of an organic unity is not adequate to these four types of organic unity, it is also not clear that it is adequate to those forms of organic unity that are represented by the *bonum variationis* and the *bonum progressionis*. Let us recall Chisholm's illustration of the *bonum variationis*. Suppose that *A* is a beautiful painting, that *B* is a painting exactly like *A*, and that *C* is a beautiful piece of music. Assume that the aesthetic contemplation of *A* has the same value as that of *B* and the same as that of *C*. The whole that is the aesthetic contemplation of *A* followed by the aesthetic contemplation of *C* is intrinsically better than that whole that is the aesthetic contemplation of *A* followed by the aesthetic contemplation of *B*. Chisholm writes, "Hence one could say that the value of a *bonum variationis* is greater than the sum of the values of its constituent parts."[5] If this varied whole is an organic unity, then it does not seem that (D5) captures it. For it would seem to be an unmixed whole that lacks a good part that is better than it.

I believe (D5) faces similar problems from the *bonum progressionis*, but I shall not pursue them here. Instead, I wish to point out that certain instances of the *bonum variationis* represent forms of organic unity that are not among Chisholm's ten types. Suppose, as before, that the aesthetic contemplation of *A* has the same intrinsic value as the aesthetic contemplation of *B* and the same intrinsic value as having a certain intellectual insight. Now suppose that *S*'s aesthetic contemplation of *A* and *S*'s having that intellectual insight is intrinsically better than *S*'s aesthetic contemplation of *A* and *S*'s aesthetic contemplation of *B*. If the former whole is an organic unity, a kind of *bonum variationis*, then it does not seem to be among Chisholm's ten types. Consider the state of affairs *S's having an aesthetic contemplation of something and S's having an insight.* If this is an organic unity, it is not a case of a defeated good, since none of the good parts are better than the whole. It is not a case of enhancement, since it is false that there is a good part of

---

[5] Chisholm, *Brentano and Intrinsic Value,* p. 71 [* p. 306 of this volume].

that whole such that every other good part of that whole has a part in common with it. If this is right, then there are more types of organic unity than the ten Chisholm distinguishes. A similar example might involve *S*'s having an aesthetic contemplation of something and someone else, *T*, having an insight. Presumably this is better than if both *S* and *T* have the same aesthetic experience or both have the same insight. Reflection on the *bonum progressionis* will, I believe, reveal similar examples.

CHAPTER 27

J. DANCY

THE PARTICULARIST'S PROGRESS*

Many people seem willing to call themselves feminists with little idea of what they are committing themselves to. The same, in my experience, is true of particularism in the theory of moral reasons. There is a common suggestion that to be a particularist is, at the outset, only to admit that circumstances can make a difference. But if that were all that particularism amounted to, it would be uncontentious. In this chapter I lay out what I think one commits oneself to if one accepts the general claim that reasons are sensitive to context – a claim sometimes called holism in the theory of reasons, and of which moral particularism is merely one expression.

Of course holism here, as elsewhere, does come in degrees. The strongest form of context-sensitivity would be the claim that every reason is somehow altered with every change of context. The weakest form is the claim that some reasons are on occasions capable of being altered by a change in context. The form of holism that I recommend is pretty weak on *this* scale, so far as the modality goes, but strong on the extent of the domain. I maintain that *all* reasons are *capable* of being altered by changes in context – that there are none whose nature as reasons is necessarily immune to changes elsewhere.

When I talk of altering a reason, I mean to suggest not that the consideration which is a reason is altered, but that its nature as a reason changes. Instead of being a reason in favour of some course of action, it ceases to be a reason for action at all, or even becomes a reason against. One could express this by saying that the practical relevance of the consideration at issue is sensitive to changes in context, and the practical relevance of the consideration includes its *polarity.* A consideration reverses its polarity when, having been a reason in favour of action, it becomes a reason against, or vice versa. My holism holds that every consideration is capable of having its practical polarity reversed by changes in context.

It is hard to be sure quite how extreme a claim this is, partly because of the awkward modality in its characterization. But I shall not be discussing that matter

* I am grateful to Eve Garrard and David McNaughton for many discussions of the issues discussed in this paper; also to Roger Crisp, Brad Hooker, Derek Parfit and to all those audiences around the world that have let me try to persuade them of the merits of particularism. I owe special thanks to All Souls College, Oxford, where as a Visiting Fellow I wrote an early draft of this paper.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 325-347.
© 2005 *Springer. Printed in the Netherlands.*

much here. Perhaps I will have to admit that not all reasons are sensitive to context in this way – that there are a privileged few, including probably the intentional inflicting of undeserved pain, which necessarily constitute the same sort of reason wherever they occur. If so, I will have lost a battle but won the war. For the main aim of my particularist position is to break the stranglehold of a certain conception of how moral reasons function – the *generalist* conception under which what is a moral reason in one situation is necessarily the same reason wherever it occurs. Generalism need not be false of *every* moral reason in order to be largely false, and hence false as a general account of moral reasons and the way they work. And if it is false as a general account of such reasons, rational constraints on moral thought and action – in particular, accounts of what consistency requires in these areas – must not themselves be based on generalist assumptions.

It may be that my train of thought here is vitiated by being overconcerned with one specific opposing account of how reasons function – Ross's theory of prima facie reasons. But I might as well admit that I do have this theory constantly in mind, since it seems to capture so well the outlines of the position I am trying to dislodge. Omitting Ross's epistemology for the moment, the theory of prima facie reasons holds:

1. What is a reason in one case is the same reason in all.[1]
2. Judgement is the attempt to determine the balance of reasons, so conceived.

1. HOLISM IN THE THEORY OF REASONS

In this section I argue in favour of particularism in ethics. In the past I tended to argue largely from example.[2] This persuades some people but not others. Here my argument will be more theoretically grounded – though there will still be considerable use of examples as well.

As I said in the preamble, I see ethical particularism as merely one expression of an overall holism in the theory of normative reasons – that is, in the theory that discusses the reasons that favour one thing (action, belief) over another. Such an overall holism can be expressed as follows:

1. What is a reason in one situation may alter or lose its polarity in another.
2. The way in which the reasons here present combine with each other is not necessarily determinable in any simply additive way.

[1] This is not fair to Ross: see D. McNaughton's 'An Unconnected Heap of Duties?' *Philosophical Quarterly,* 46 (1996), 433-47.
[2] See my *Moral Reasons* (Oxford: Blackwell, 1993), ch. 4.

There are theoretical reasons and practical reasons, reasons for belief and reasons for action. My holism is intended to hold on both sides of that distinction. I start by trying to establish that theoretical reasons are holistic. We will quickly find that theoretical reasons are perfectly capable of changing their polarity according to context, without anyone making the slightest fuss about the matter. For instance, suppose that it currently seems to me that something before me is red. Normally, one might say, that is a reason (*some* reason, that is, not necessarily sufficient reason) for me to believe that there is something red before me. But in a case where I also believe that I have recently taken a drug that makes blue things look red and red things look blue, the appearance of a red-looking thing before me is reason for me to believe that there is a blue, not a red, thing before me. It is not as if it is some reason for me to believe that there is something red before me, but that as such a reason it is overwhelmed by contrary reasons. It is no longer *any reason at all* to believe that there is something red before me; indeed it is a reason for believing the opposite.

As I say, it seems to me that nobody ever thought of denying what I am claiming here. I know of nobody who has nailed themselves to an atomistic (i.e. non-holistic) conception of how theoretical reasons function. If generalism is taken to be the view that all reasons are general reasons, i.e. that if a feature is a reason in one case, it is the same reason in any other case, generalism is uncontentiously false of theoretical reasons.

Let us now turn to ordinary practical reasons. We will find just the same thing there. There are plenty of examples to persuade us that such reasons are holistic (or non-generalist, if you like). For instance, that there will be nobody much else there is sometimes a good reason for going there, and sometimes a very good reason for staying away. That one of the candidates wants the job very much indeed is sometimes a reason for giving it to her and sometimes a reason for doing the opposite. And so on. Now examples would be of little use if there were some theoretical obstacle to taking them at face value. But again we should remind ourselves that nobody has ever really debated the question whether ordinary practical reasons are holistic or not. There should be no *parti pris* on this issue; so the examples, which are legion, should be allowed to carry the day without resistance.

Perhaps this is too quick. There is a theory-based reason for doubting my claim that practical reasons are holistic, one that derives from the common thought that practical reasons are grounded in desires of the agent in a way that theoretical reasons are not. What one wants should not affect what one judges to be the case, on pain of charges of bias or prejudice. But what one wants can perfectly well affect what one has reason to do. Indeed, many find it hard to conceive of our having any practical reasons at all if we had no desires. My own view on this matter, however, is that desires do not give us or ground our reasons. Reasons stem from the prospect of some good. If we have no other reason to do a certain action, wanting to do it will give us no reason at all; nor can wanting to do a silly action make it marginally less

silly. (These are only the first moves in a long debate.[3] I mention them here only to show the sort of way in which I find myself denying the possibility of grounding practical reasons in desires of the agent.) This view of mine is, of course, an independent input in the present debate. I mention it only to show that a certain motive for doubting the analogy I have been drawing between theoretical and practical reason is itself contentious.

It may be that here we come across the real motivation for generalism in the theory of practical reason – an adherence to the view that reasons for action are partly grounded in desires. For if we accept that view, and if we then think of desires as giving the desirer the same reason wherever the desire occurs, we will at least get the sort of generalism I discussed above. The right response to this, however, is to claim that even if all practical reasons are grounded in desires, the same desire need not always function as the same reason. Consider first the third-person case. That he wants power and she does not may be a reason to give the power to her rather than to him, as I have already said. (It may at the same time be a reason to give it to him, since according to me one feature can be a reason on both sides at once; but remember that here it is a reason not to give it to him, and that it need not always be such a reason.) Now consider the first-person case. Suppose that I am trying to train myself into indifference towards a girl. I want very much to spend time with her. But I also want not to have this want, since she is permanently indifferent to me. It is better for me not to think of her at all. If I spend time with her, this will make things worse for me rather than better – so long as I have not yet succeeded in training myself into indifference towards her. Once I am indifferent towards her, I can spend time with her without loss. In this situation, it seems, my desire to spend time with her may be a reason for me not to do so.

Before carrying on to consider moral reasons, which have been claimed to be non-holistic, I want to step aside for a moment to ask whether I have not already made a mistake. There is a distinction between epistemic and what one might call constitutive reasons. An epistemic reason is a reason for believing something or other; a constitutive reason is a reason why something or other is the case. That the butler's fingerprints are on the murder weapon is a reason for believing that he did the deed, but no part of what makes it the case that he did it or of why it is true that he did it. That the hedgehogs are hibernating early is a reason for believing that we will have a severe winter, but not any part of what makes it the case that the winter will be severe. And so on. Now holism in the theory of reasons should concern itself with constitutive reasons rather than with epistemic ones. But I appear to have argued only that epistemic reasons are holistic, for my first example, or fulcrum, concerned reasons to believe that there was a red thing before me. It is, therefore, technically irrelevant.[4]

This is true, and I apologize for it. But matters can be redeemed. We should not suppose that all that I have shown is that epistemic reasons are holistic, it being left

[3] For the remainder of the debate, see ch. 2 of my *Practical Reality* (Oxford: Clarendon Press, 2000).
[4] Thanks to Nick Zangwill for pointing this out to me.

entirely open whether constitutive reasons are or are not. For many, possibly most, epistemic reasons are also constitutive. For instance, that an action involves the gratuitous inflicting of pain is held by many to make it wrong, but equally clearly functions as a reason to believe that the action is wrong. It is both an epistemic and a constitutive reason. Some epistemic reasons are not constitutive, and *perhaps* some constitutive ones are not epistemic; this is all that can be said. Now could it be the case that the epistemic ones are holistic but the constitutive ones are not? I think that this is inconceivable. The mere fact of the overlap between reasons of the two sorts should give us pause. But more importantly, can we suppose that the very logic of epistemic reasons is capable of differing at a very deep level from that of constitutive reasons? This supposition entirely undermines the sort of connection there needs to be between reasons why things are so and reasons for taking them to be so.

I return, therefore, to the onward or outward spread of holism. So far we have it that theoretical reasons (constitutive ones) are holistic, and so are ordinary practical ones. Now could it be the case that moral reasons are quite different from others in this respect, being the only atomistic ones? This is what many have supposed, in supposing that moral rationality is based on the existence of a range of moral principles. Moral reasons, they have held, necessarily behave in regular (or rule-bound) ways, though other reasons see no need to behave in that way at all. About this I want to say that straight off it just seems incredible that the very logic of moral reasons should be so different from that of others in this sort of way. Consider here the sad fact that nobody knows how to distinguish moral from other reasons; every attempt has failed. How does that fit the suggestion that there is this deep difference between them? Not very well at all. Then of course there are examples to be considered, examples of apparently moral reasons functioning in a holistic way. I forbear to bore you with these. It just seems inevitable that moral reasons should function holistically in the way that other reasons do.

This certainly makes it hard to hold, as many do, that the very possibility of moral distinctions, of moral thought and judgement, is predicated on the existence of a range of moral principles. Moral principles, however we conceive of them, seem all to be in the business of specifying features as *general* reasons. The principle that it is wrong to lie, for instance, presumably claims that mendacity is always a wrong-making feature wherever it occurs (*pro tanto,* of course, not necessarily absolutely). It cannot be merely a generalization, a claim that lies are mostly the worse for being lies, for if all moral principles were of this sort the argument that moral thought and judgement depends on the possibility of moral principles would simply be the argument that such thought is impossible unless there is a considerable preponderance of normal cases over abnormal ones. I have never seen this argument made, and I doubt, what is more, whether it would be persuasive if restricted to ethics.

If moral reasons, like others, function holistically, it cannot be the case that the possibility of such reasons rests on the existence of principles that specify morally

relevant features as functioning atomistically. A principle-based approach to ethics is inconsistent with the holism of reasons.

All the same, it might be argued, we have to admit that there are some invariant reasons – some features whose practical relevance is invariant. And surely I should allow this, because holism, as I expressed it, concerns only what may happen, not what must. It could be true that every reason may alter or lose its polarity from case to case, even though there are some reasons that do not do this. If they don't do it, this will be because of the particular reasons they are. Invariant reasons, should there be any, will be invariant not because they are reasons but because of their specific content. And this is something that the particularist, it seems, should admit. It is like the claim that a man can run a mile in four minutes, that Sam Smith is a man, and Sam Smith cannot run a mile in four minutes. These claims are compatible, and so are the claims that reasons are variable *qua reasons* though some reasons are (necessarily, given their content) invariant. The invariance, where it occurs, derives not from the fact that we are dealing here with a reason but from the particular content of that reason.

So can the particularist admit the existence of *some* invariant reasons? The obvious examples are things like the causing of gratuitous pain on unwilling victims. Surely, it is commonly urged, this is always for the worse, even if over all we might in some case be morally forced to do it. Well, the first thing to say is that admitting the possibility of some invariant reasons is a far cry from admitting that the very possibility of moral thought and judgement is dependent on our being able to find some such reasons. To support any such suggestion, we would somehow need to be able to locate a sufficient range of invariant reasons, ones that together somehow covered the moral ground entirely and themselves explained the nature and role of the variant reasons. This is quite a different matter from simply trying to refute particularism (which is merely an application of holism in the general theory of reasons to the moral case) by producing one counter-example of an invariant reason, which is normally what is going on.

Further, we should remember that the question whether reasons are atomistic or holistic is a very basic question about the nature of rationality, of how reasons function from case to case. It is, I suppose, conceivable that though the vast bulk of reasons function holistically, there are a few that function atomistically. But if this were true we would have a hybrid conception of rationality. There would just be two sorts of reasons, each with their own logic, and moral thought would be the uncomfortable attempt to rub such reasons together. It is *much* more attractive, if at all possible, to think of our reasons as sharing a basic logic, so that all are atomistic, or all holistic.

Let us consider, then, how the supposed invariant reasons function as reasons in the particular case. Take the well-known example of the fat man stuck in the only outlet from a cave that is rapidly filling with water from below. We and our families are caught in between the fat man and the rising water. But we have some dynamite. We could blow the fat man up and get out to safety. But the fat man is unwilling to be blown up (he, at least, is safe from drowning); and, let us immediately admit, he

is blameless in being where he is, and in being fatter than the rest of us. So what we propose to do involves the destruction of an unwilling and blameless victim. As such, we might say, this is *some* reason against lighting the fuse and standing back. The question I want to raise is whether this feature (that we are causing the death of an unwilling and blameless victim) is functioning as the reason it here is, in any way that is to be explained by appeal to the (supposed) fact that it functions in the same way in every case in which it occurs at all. It seems to me that this feature is the reason it is here quite independently of how it functions elsewhere.

Of course if the feature is genuinely an invariant reason, this fact, should we discern it, will be of use to us in any case where we might be in doubt as to the contribution it is making. We can say 'This is an invariant reason, it makes such-and-such a difference there, and so it must be making that difference here.' But suppose that we were to treat one of these supposedly invariant reasons as potentially variant, so as to deny ourselves the use of that inference. What sort of mistake would we have made? Would it be a failure of *rationality* to treat an invariant reason as potentially variant, or just a mistake of fact? I suggest that the invariance of the reason is an epistemic matter rather than a constitutive one. That the reason functions invariantly is a clue to how it is functioning here, but in no way constitutes the sort of contribution it makes to the store of reasons here present. In that sense, the invariance of its contributions is not a matter of the logic of such a reason, and failure to treat the reason as functioning invariantly is not a failure to understand how it functions *as a reason.* It is a perfectly good reason case by case without our worrying about how it operates elsewhere.

I conclude, then, that particularism should accept the possibility of invariant reasons, so long as the invariance is not a matter of the logic of such reasons, but more the rather peculiar fact that some reasons happen to contribute in ways that are not affected by other features. We can admit this without adopting a hybrid theory of rationality, so long as we treat the invariance of any invariant reason as an epistemic matter rather than as a constitutive one.

2. HOLISM IN THE THEORY OF VALUE

The next question concerns whether our holism in the theory of reasons spills over to generate a holism about *value.* This new holism, value holism, can come in various forms, just like the holism of reasons. In broadest outline, in my hands it will amount to the claim that for any x that has value in some context, x may have a different value or none at all in other contexts, and if there is disvalue as well as value, x may have value in some contexts and disvalue in others.

Is there any prospect of accepting a holism of reasons and denying value holism? This is not a matter on which there is a long history of debate. Presumably we should approach the issue by thinking about the relation between reasons and values. Thomas Scanlon has recently revived the view that value is to be understood in

terms of reasons.[5] He defines a valuable object as one that has features that give us reasons to protect, promote, admire, respect, approve of it (etc.: this list is open-ended). On this view, it seems inevitable that a holism of reasons will generate a holism of value. There are, however, other views to be considered. One might allow that wherever there is value, there are reasons, but leave it open whether there can be reasons that are not directly connected with value.[6] In this way one would make room for certain forms of deontology, without moving very far from the Scanlon position. Michael Slote and Roger Crisp suggest that there are some reasons that do not derive from values.[7] One might also ask whether there are agent-relative values as well as agent-relative reasons; perhaps agent-relative reasons do not 'stem from' values at all. There is, then, a spectrum of views to be considered. On the most trenchant views, a holism of reasons must be matched by a holism of values. But even on the less trenchant views, it still remains possible that the holism of reasons must match that of values. Suppose we agree for a moment that *most* reasons are linked to values. ('Linked to' is the vaguest phrase I can think of.) And suppose that values are atomistic, i.e. insensitive to context. Immediately we have the problem of explaining how it can be that the reasons linked to those atomistic values are able to be holistic at all. Why are they not atomistic too? Of course, if there are some reasons that are not linked to values, and those too are holistic, we might think this some justification for supposing that the explanation of their holism will not make reference to any holism of values. And we could then argue that, by parity of reasoning, the holism of other reasons is to be explained in a similar non-value-related way. One trouble with this procedure, however, is that the preponderance of value-linked reasons seems to be enormous, at the least, and we are therefore in danger of letting the tail wag the dog. Another trouble with it is that the connection between reasons-holism and value-holism seems so plausible. It is far easier to explain the holism of the few non-value-linked reasons in terms of their relation to the holism of the many value-linked ones than to cast off the only obvious prop we have.

I take it, then, that reasons-holism does not entail value-holism, since it is at least possible that reasons-holism is to be explained in other ways. By far the most plausible picture is that, just as most reasons are linked to values, so their holism is linked to a holism of values. This leaves us with a strong incentive to be particularists in the theory of value. Of course this incentive needs to be supported by examples of values varying according to context. But these are not too hard to produce. The disvalue of pain may be affected by the question whether it is part of a

---

[5] See Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1999), 95-97. Similar views were considered by W. D. Ross and adopted by A. C. Ewing. See Ross's *Foundations of Ethics* (Oxford: Clarendon Press, 1939), 278-83, and Ewing's *The Definition of Good* (London: Macmillan, 1947), 148-49.

[6] I consider this possibility in my 'Should We Pass the Buck?', in A. O'Hear (ed.), *The Good, the True and the Beautiful* (Cambridge: CUP, 2000) [* pp. 33-44 of this volume].

[7] See M. Slote, *Goods and Virtues* (Oxford: Clarendon Press, 1983), ch. 5, and R. Crisp 'The Dualism of Practical Reason', *Proceedings of the Aristotelian Society* 96 (1996), 53-73.

merited punishment. This does not mean that the pain hurts less, but that the punishment is not as much the worse for involving the infliction of the pain as we might have been led to think by considering the disvalue of other pains that hurt as much.

If value-holism shadows reasons-holism, the two views have to be structurally similar. Now as far as reasons go, the holism that appeals to me holds:

1. What is a reason in one situation may not be the same reason in another; it may even change its polarity.
2. The way in which the reasons here present combine with each other is not necessarily determinable in any simply additive way.

By analogy, then, our value-holism should look like this:

1. A feature or part may have one value in one context and a different or opposite value in another.
2. The value of a complex or whole is not necessarily identical with the sum of the values of its elements or parts.

And this is therefore the form of value-holism that I adopt. In this I differ from G. E. Moore, despite the fact that we could both be called 'organicists' in the theory of value. Moore believed (2) above but not (1). He held that any feature or element necessarily retained the same value as it moved from context to context, but that it could contribute to a complex of which it was a part a value other than the one that it had there. The whole, that is, could be more valuable because of the presence of a certain part than could be explained by the value of that part; a part can contribute more, or less, value than it actually has.[8] I don't believe any of this.

What explains the difference between Moore and me? This difference needs explaining, since all the examples that impressed Moore are just the sort of examples that impress me. How then have we come to such different conclusions? The answer lies in the fact that Moore accepted without question a certain doctrine of supervenience. He believed that the intrinsic value of something supervenes upon its other intrinsic qualities, and so that where an element does not change its intrinsic qualities on moving from one complex or whole to another, it must retain the same intrinsic value. The examples of organic wholes that impressed him required him therefore to say that elements can contribute to a complex more, or less, value than they have got themselves there. I, however, do not accept Moore's doctrine of supervenience. I accept (or used to accept, and as far as the present debate goes continue to accept) a slightly but crucially different supervenience doctrine, that intrinsic value supervenes upon other qualities, but not that the intrinsic value of one

---

[8] See G. E. Moore, *Principia Ethica* (Cambridge: CUP, 1903), 30.

object supervenes upon *its* other qualities. For me, it supervenes upon other qualities, including those of other objects.[9]

This may seem perverse. But there is a ready explanation of it, once one remembers my commitment to particularism. First, I distinguish between those features from which some value *results* (the good-making features, as we might put it), and other features whose presence or absence would have made a difference. The latter features are obviously relevant to the value, but they are not playing the same role as that played by the good-making features; they are not themselves part of what we might call the 'resultance base'. Given this distinction between roles, I can announce that intrinsic value is value that results from intrinsic properties of the object concerned, but also allow that that value can vary because of changes elsewhere, that is, in those properties whose presence or absence can make a difference to the ability of the intrinsic properties to generate the value that they do. The notion of supervenience draws less fine distinctions than that of resultance, and all we can say is that intrinsic value supervenes upon other properties, not particularly upon intrinsic properties. And once we have said that, there is no reason to stop short of allowing that the 'other properties' can be properties of other objects, or relations between them and the first one, or whatever.

My doctrine, then, is a sort of global supervenience, since the supervenience base is cast so wide, while Moore's is a sort of local supervenience. Who is right? I do not see that there is any way of determining the answer to this question before we get down to arguing about particular cases. It is not, that is, going to be a logical question which of us is right. Nor is it going to be decided quickly by appeal to the notion of a reason, or to naturalism in metaphysics, or anything like that.

So the overall situation, as far as value-holism goes, is that if values are to track reasons, and if the structure of value-holism is therefore to be the same as that of reasons-holism, we have to abandon one traditional formulation of the doctrine of supervenience in favour of something less familiar. But to appeal to the traditional doctrine to defeat my form of value-holism would be to beg the question.[10]

3. HOLISM IN THE THEORY OF CHOICE

Let us suppose now that I am right on both counts: my reasons-holism is the truth, and my value-holism is the truth as well. What does this tell us about the possibility of a full ordering, in which everything has its place, and where for each A and each B, A is either better than B, worse than B, or roughly as good as B? Well, so far as what we have so far seen will take us, there might still be a full ordering of that sort,

---

[9] For an early proponent of this view of supervenience, see A. C. Ewing, *The Morality of Punishment* (London: Kegan Paul, 1929), 166: 'It does follow from the conception of goodness or value that the value of something cannot be different except as the result of some other difference, but this difference need not necessarily lie in the thing itself, it may lie in something else. We cannot therefore say that the intrinsic value of any quality will always be the same under all circumstances.'

[10] It seems to me that Susan Hurley does make this appeal in her *Natural Reasons* (Oxford: OUP, 1989), 235-37.

in which everything has its place in the table of values, from best to worst. For all the considerations we have so far adduced concern the way in which the value of a complex is determined from the values of its parts, and the way in which parts may change their values as they move from one complex (context) to another. Once the value of the whole is determined, however, it is not going to vary, and we can enter the whole in its proper place in the great ordering. This would mean, for instance, that transitivity is not threatened by particularism. For with everything in its own place, we are never going to get a situation in which A is above B, B above C, and C above A in the Great Order.

This pleasing picture is however not as secure as it initially appears, sad to say. Let us remember that some of the things that have value are actions, and actions are chosen out of a set of alternatives available to the agent at the time. Now to adopt the picture I have just described is to suppose that each alternative action has its place in the great ordering of values, a place that is not affected by the place or nature of other alternatives. And this seems to require that the value of an action is never affected by the question what the available alternatives are. Now this is a very attractive doctrine indeed, partly because it enables us to retain a plausible principle of rational choice which we might call the principle of the indifference of independent alternatives (IIA):

> IIA: If in one situation I prefer action A to action B, it can never be rational for me to prefer B to A in other situations which differ from the first merely in the fact that further alternatives are available.

In simpler terms, if I choose A where my choice is between A and B, I cannot rationally choose B where my choice is between A, B and C. The availability of C may indeed alter my overall choice, but it cannot affect the relative ranking order of A and B that has already been established.

Though this principle is *very* attractive, I am not convinced that stubborn adherence to it is fully compatible with the broadening particularist perspective that I have been developing. For it is not obvious to me how one can prevent available alternatives from counting as part of the context within which an action is placed. And if one cannot do this, then the general particularist claim that context can make a difference to value as well as to reasons seems likely to take us to the view that the value of an action or choice can be affected by the alternatives that are available at the time.

There is a reply to this, however. The argument of the previous paragraph might have been merely that every alternative is an object, though not all objects are alternatives (to each other). Since every object may have its value affected by others, every alternative may have its value affected by other objects, including some that are alternatives to it. There could be no bar against this happening – no bar that ensured that only those objects that are not alternatives to this one are capable of affecting its value. But this, though true, does nothing to establish the controversial

doctrine that is really what we are after. That doctrine is that when one object becomes an alternative to another, that change may make a *further* difference to the value of the second – a difference beyond that made by the existence and nature of the first object. And this doctrine does seem very peculiar.[11]

But not as peculiar as all that, perhaps. There may be examples of this phenomenon – of the arrival of a new alternative making a difference to the relative values of two existing alternatives, in ways that are not explained merely by the *existence* or possibility of the thing that becomes an alternative, but rather by its new status as an alternative. Suppose that we have two alternatives A and B, and that we prefer B to A. Our original question was whether we might be rationally led to prefer A to B if there appears a further alternative, C.[12] Any successful example of this must meet certain criteria. It must not be one in which we simply change our minds about our initial ranking of A and B, perhaps for the reason that the appearance of C as an alternative draws our attention to something that we had previously missed. If nothing changes, however, it is hard to see how the ranking of A and B can be reversed by the arrival of C. The question to bear in mind is whether the examples offered contain the right or the wrong sort of change.

Suppose that I have to buy a house in Reading, and have a choice between a smaller house within walking distance of the university, and a larger and more expensive one that requires a bus ride. I prefer the larger one despite the bus ride. Then a third house, even larger but also further away than the second, comes onto the market. I realize that if buy the second house, I will always regret not having bought the third. With this in mind, I buy the first house. Is this rational? I suggest that it can be.

There is of course a change of information here, but it is not the *sort* of change that I tried to rule out earlier – a change that leads to a change in my initial ranking. If there were no further house available, I would still have chosen the second. I have not changed my mind about *that.* So we are not dealing here with a simple revision of the initial choice in the light of new information.

Have I cheated in the description of the example? The obvious mistake would be a slide in the nature of the 'objects' of choice. Let us be sure that these 'objects' are my buying house A, my buying house B and my buying house C. House C was there all the time, but it was not on the market. The mere possibility of a further house C that is even larger than B, though still affordable, is not enough to cause me to prefer

---

[11] I have expressed this doctrine in terms of a *change.* But that need not be the point. The question could equally well be phrased in terms of the difference between the case where C is not an alternative and the case where it is; can the difference between C's being an alternative and its not being one make a difference to the relative values of A and B? Here there is no talk of change. There is, of course, nothing wrong with examples that do involve change. It is just that I should be careful to avoid supposing that change is essential to the point.

[12] It is not, of course, strictly necessary for us to find an example in which the order of the initial choice is reversed. It would be enough if we found a case in which the relative values are altered, so that the one that we originally preferred we still prefer, but not by so much – or by more, perhaps. Then we could argue that this sort of change in relative preference is bound to lead, on occasion, to a change in ordering. But it is more striking to produce an example in which the ordering is reversed.

the nearer house. There has actually to be such a house C available to me before my continual regret at not having bought it can turn into a reason for me to choose the nearer one.

The simplest example I know, and perhaps for that reason the best, is found in Wilfred Thesiger's *Arabian Sands*.[13] He wrote, about his travels across the desert on a camel: 'I would not myself have wished to cross the Empty Quarter in a car. Luckily this was impossible when I did my journeys, for to have done the journey on a camel when I could have done it in a car would have turned the venture into a stunt.'

So it does look as if there may be examples in which independent alternatives are not indifferent.[14] Maybe, then, even if particularism does commit us to the existence of such examples, this is not a disaster. But I raised a question earlier that I have not yet answered. Does particularism itself constitute a reason for rejecting IIA? My first attempt to show that it does was a failure. I argued that particularists should not be surprised to see the nature of one thing making a difference to the value of another. But this was irrelevant. The real question was whether the rather special feature of 'being an available alternative' can make a difference to the relative values of two things. Particularism does not show that this must happen. Particularism is everywhere permissive rather than prescriptive; or perhaps we should say that it forbids some things and prescribes nothing but suspicion. In the present case it says that we should be open to the possibility of such a thing and not make a fuss if some crop up. The only reason for supposing that there *cannot* be any examples would be the generalist claim that since the feature of 'being an available alternative' often makes no difference, it makes no difference anywhere.[15]

---

[13] London: Longmans (1959), 260.

[14] I am grateful to Lars Bergström for very helpful discussion of this and other potential examples. He referred me to M. Resnik, *Choices* (University of Minnesota Press, 1993), 40, for another example with the same structure. Here is a further one, which I owe to Eve Garrard. I have to choose between two men, Joe and Sam. Joe is dull but reliable, Sam is unreliable but exciting. I prefer Sam, because if I chose Joe I would always be missing the excitement that Sam would have given me. But then Sebastian comes along, who is even more exciting but yet more unreliable. I realize that if I chose Sam, I would always be missing Sebastian's excitements, even though Sebastian's unreliability is so terminal that he is ineligible. And this tells me that my reason for preferring Sam to Joe is no reason, in the new situation, since choosing Sam will not lead to my having no lost excitement to regret. So I choose Joe.

[15] An interestingly different avenue of approach, which I will not pursue here, starts from something that Derek Parfit is apparently happy to admit, namely an analogous claim concerning not value but 'ought'. Parfit's view seems to be that it is possible that one ought to do A if B is the only alternative, but that if C is also available one ought to do B. He denies, however, the claim that the *values* of A, B and C can be related in a structurally similar way. See his *Reasons and Persons* (Oxford: OUP, 1984), 429: 'Whether I ought to act in one of two ways may depend on whether it would be possible for me to act in some third way… I then ask whether, compared with A+, A would have been better. The relative goodness of these two outcomes cannot depend on whether a third outcome, that will never happen, might have happened.' So in the evaluative realm Parfit is what one might call a choice-atomist but in the deontic realm he is a choice-holist. My own view is that this position is unstable. And my reason is fairly predictable: that if one admits that there are examples of choice-holism in the deontic realm, structurally similar examples will emerge in the evaluative one.

There remains a difficulty. There appears to be an argument that there could be no counter-example. If so, the situation is unstable. We would have an example, our holistic position that there could be examples, and an argument that there can't be. Here is the argument. Sadly, it is one to which there can be no particularist objection *as such*; it is not a switching argument, for instance.

Let us start with a supposed overall ranking, the Great Ordering. Everything has its place in the order. We can compare the values of different objects, and compare different objects in order to establish their relative placing, and these activities cannot themselves alter the values of the objects compared or their relative places; otherwise the very notion of relative value would be incoherent. One can compare the values of merely possible objects, e.g. possible courses of action or possible states of affairs. Could there be any difference between the values of merely possible objects and the values of those objects should they become real? No; for otherwise the activity of establishing the relative values of different possibilities would be incoherent. And this would make deliberation before action incoherent, if deliberation is the establishing of relative values of possibilities so as to decide which to make actual. Suppose then that I ask you to rank ownership of each of ten paintings. What you are ranking is a set of possibilities. Suppose then that I give you all the paintings, and ask you to rank the ten actual ownings. There can be no conceivable reason for a change in your ranking order (unless of course you have changed your mind). Now: could there be a difference between an order of preference and a ranking order for choice, where what one is dealing with is alternatives? No: there is no possible relevant difference between a preference order and a choice order. Suppose that instead of giving you all the paintings, I give you the money to buy one. You should buy the one that came top in your preference ranking. So the feature of 'being an alternative' cannot make a difference.

Matters are more complicated than this. The complications do not make a difference, but they are relevant to what happens later. There is an obvious difference between buying just one painting and ranking them all. The difference comes out when we consider a case where I give you the money to buy one painting, then enough to buy another, and so on until you have all ten (though you never knew in advance that I would give you the money for the next). The order in which you buy the paintings need not be the same as your original order of preference. To see this point, it is important to distinguish between two quite different preference orders. The first has ten slots, in each of which one is asked to put one item of the form 'I own picture n'. The second has ten slots, the first of the form 'I own picture n', the second of the form 'I own pictures n and m', the third of the form 'I own pictures n, m and p', and so on. There is no reason whatever why either of these two orders should be extractable from the other. The point is that if you already have, say, six of the paintings, you might rationally choose to add to those six a painting other than the one that came seventh on the list. To get a true analogue of the original ordering, when it comes to choice, we have to suppose that I give you enough money to buy one, but that just as you try to buy it someone else gets in first; so you should go for the second on your original list – but the same happens

again, and so on down to the tenth. The order of choice should be the order of preference.

What we have, then, is an explanation of why a certain feature, 'being an alternative', cannot make a difference, and therefore of why particularism is compatible with a full ordering. In one sense (epistemically, perhaps) it is *possible* that 'being an alternative' can make a difference. But there is an argument (which is not a switching argument) that no instance of this could be found. So the situation seems to be that we have on the one side an example in which the feature does make a difference, and a weak general reason derived from our holism to expect this sort of thing to crop up, and an argument on the other side to show that it is impossible. Now this is not one of those situations in which there can be reasons on one side and reasons on the other, and we can just decide where the balance of probability lies, leaving the defeated reasons in place. If we go with the example, we have to show what is wrong with the argument on the other side.

Luckily this can be done. The property of being an alternative is incapable of making a difference to a ranking order already established because there is no relevant difference between overall preference and overall choice. And just as the ranking preference order may be affected by the list of things to be ranked, so that if we take something off the list, the rankings of the rest may change, so with the ranking of alternatives. But all that this shows is that *preference is like choice*, and like choice in the crucial respect that it deals with alternatives. Being alternatives is the same as being mutually exclusive. Not all preference rankings are rankings of objects conceived as mutually exclusive, as we have seen. But some are. And the same is true whether we are ranking existing objects or possible ones. So the explanation of why the feature of being an alternative cannot be the cause of a difference between a preference ranking and a choice ranking is that this feature is present on both sides.[16] Our conclusion should be that being an alternative can make a difference to all three rankings: of possible objects, of actual ones, and of objects of choice.

The existence of persuasive examples should then move us without further resistance from value-holism to a sort of choice-holism, which holds that:

1. The value of one alternative can be affected by the nature of other available alternatives.
2. Assessing the relative merits of different alternatives is not the same as assessing the various alternatives one by one and then comparing the results.

---

[16] Thanks to Eve Garrard and David McNaughton here. It would be wrong to say, in reply to this argument, that the feature we were originally discussing was that of being an available alternative, not that of being an alternative. The notion of availability merely takes us from possible choices (preferences) between mutually exclusive options to actual choices.

I want to end this section by comparing what I have said here with something I wrote in *Moral Reasons:*

> My daughter trod on a sea-urchin on holiday a few years ago, and we caused her considerable pain (not entirely with her consent) in extracting the spines from her heel. Was the pain we caused her something which made our actions worse than they would otherwise have been? Here is a switching argument which says that it was. Had there been available a painless method of getting the spines out, we would and should have adopted it. We would have been wrong to continue digging in her heel with a needle, because of the pain. Surely this shows that as things were our actions were the worse for the pain they caused?
>
> I don't think it does show this. What we should say about cases like these is that a feature which would have made this sort of difference had there been any alternative choice need not necessarily make it if there is no alternative. It seems to me quite consistent to say that as things stood our action was not the worse for the pain it caused, though that pain should have led us to choose another method had one been available.[17]

The idea, expressed in terms of reasons rather than, as above, in terms of values, was that the pain is not a reason against the action if there was no alternative, pain-free course of action available. It is not just that it is not sufficient reason; it is not any reason at all. I presented this thought as an application of a style of switching argument, whose general form is: if this action were less F, it would be better; so its being F must detract from its overall value. But it can be seen immediately that the example I gave goes further than is required for that purpose. My use of an example that hangs on a point about alternatives was more of a distraction than a help, since the general point I was trying to make was nothing to do with alternatives. Talk about available alternatives was intended more as an explanation of the supposed fact that, in the example given, the action was not the worse for the pain caused, even if it would have been better with less pain. Since it was not possible to do it with less pain, the pain caused does not make the action worse than it would otherwise have been. In possible world terms: even though, in the nearest world in which there was an available pain-free alternative, the action we did was wrong, wrong because of the pain it caused, and the worse for that pain, the actual action is not the worse for that pain. *This is just an application of the holistic thought that a feature can make a difference in one situation that it does not make in another.* Where this occurs, particularists admit that there must be an explanation of it; the explanation is that in the actual world, there was no alternative.

Perhaps, then, the situation is like this. Holism takes away from us one of our two main reasons for sticking to principles like IIA. If one is a holist, it is going to be hard to think that the question whether something was a real alternative cannot make a difference. If IIA expresses a form of generalism, holists don't have *that* reason to believe it. They might have the other reason, which is that if we lose

[17] Dancy, *Moral Reasons*, 65-66.

principles like IIA (and all the rest), we lose what is really the only detailed account of the 'logic' of choice. The loss of IIA seems to be another nail in the coffin of the idea that there is such a logic.

One might suggest, however, that the loss of IIA is far worse for generalists than it is for particularists. For generalists, the pillars of practical rationality would really be tottering.

4. CHOICE-HOLISM AND THE POSSIBILITY OF A FULL ORDERING OF VALUES

Suppose then that we accept a general choice-holism to go with our reasons-holism and our value-holism. Have we totally lost any possibility of a full ordering? There are two ways in which we could hope to retain anything like a full ordering.

The first is to say that what we have established is only that the context of real choice (i.e. the actually available alternatives) can affect the value of an option. This result, we might say, is clearly disturbing. But it does not altogether disturb our full ordering. Choice-holism concerns itself with real situations, in which the question what alternatives are available is a serious practical one. As such, it is to be distinguished from any thoughts about the effects of *merely comparing* one option with another. The value of an option will not vary according to what we imagine as the possible alternatives to it; it will only be affected by what actually are the alternatives. And mere comparison is relevantly similar to imaginary choosing, we might say, so long as the purpose of the comparison is to establish relative value. So an object's value is not affected by the mere act of mental comparison with another; it is only able to be affected if the two objects become real alternatives for some agent.

Even this, of course, will do something to upset our full ordering, if we suppose that the very same thing can occur in more than one actual choice situation. But we might deny that possibility, supposing instead that objects of choice are incapable of recurrence. We cannot have the same action again, that is, and we cannot have any other choosable object again either. For the objects of choice are not repeatables. If I offer you a chocolate bar today and you refuse it in favour of a pint of beer, and I offer the very same bar to you tomorrow, the fact that it is the same bar does not show that you have the same choice again. You only have a similar one, and holists allow that objects that are intrinsically similar may yet differ in value because of their context.

Be that as it may, the position that this move is trying to defend is surely another unstable one. It holds that mere comparison of two objects A and B, which we can do at any time at will, is incapable of affecting the values of A and of B; but should they become actual alternatives, their values may be affected. So I may compare A and B and prefer B, and yet when I have to choose between the two choose A without irrationality. This distinction between actual and imagined choice, or between the effects of choice and those of comparison, is surely unsustainable. And this means that the dream of a full ordering collapses entirely. For if I cannot

compare two objects without being in danger of affecting their relative values by doing so, there is surely no sense left in which objects have their own place in the ranking order. The ranking order must mean that objects have their place on it whether one actually compares them or not; indeed, to compare them is just to establish their relative placings in the order. If one could affect those placings by the act of comparison, the notion of an order would be destroyed.

So much for the first way of defending the possibility of a full ordering. The second way involves us in redescribing each option in terms of the available alternatives to it.[18] Instead of thinking of ourselves as having the three options of buying house A, buying house B, and buying house C, our three options are:

1. Buying house A when we could have bought houses B or C.
2. Buying house B when we could have bought houses A or C.
3. Buying house C when we could have bought houses A or B.

Now, we might say, these three options have an unvarying value, and occupy a fixed place in the full ordering. For if one of the three houses is taken off the market, or a fourth house enters the equation, we no longer have any of these three options, but either two or four new ones.

The difficulty that I see in this approach is that if it is to avoid the difficulty we have already exposed, that comparison of the value of two objects is relevantly similar to choice, we will have to relativize every item on the full ordering to all other items, first severally and then in pairs and so on up until each is relativized to all others at once. And there will be no way of predicting, from the value of an option that is relativized to degree n, what its value will be relativized to other degrees. Given this, the use of the full ordering will be limited indeed. Transitivity, for instance, must fail. For if we rank ‘A when we could have had B’ above ‘B when we could have had A’, and ‘B when we could have had C’ above ‘C when we could have had B’, it in no way follows that we should rank ‘A when we could have had C’ above ‘C when we could have had A’.[19] It will be perfectly true, that is, that every relativized option occupies one and only one place in the ordering, without this doing anything to preserve the conception of rational choice that the ideal of a full ordering was designed to promote.

There is a further problem. Suppose that we have a full ordering of all relativized options. This locates each option with respect to every other. Suppose now that I ask of item 32 in the list how it compares in value with item 33. It need not be the case

---

[18] I have borrowed this manoeuvre from John Broome, *Weighing Goods* (Oxford: Blackwell, 1991), ch. 5, and ‘Can a Humean be Moderate?’ in R. G. Frey and C. Morris (eds.), *Value, Welfare and Morality* (Cambridge: CUP, 1993), 51-73.

[19] Of course, relative to one and the same three-way comparison, transitivity must be preserved – or at least nothing that I have said gives us any reason to dispute that. If we do dispute it, we will probably do so for quite different reasons, i.e. those stemming from comparisons in which many different criteria are operating at once. For a recent rehearsal of such considerations, see Larry Temkin, ‘Rethinking the Good’ in J. Dancy (ed.), *Reading Parfit* (Oxford: Blackwell, 1997), 290-345.

that my answer is that item 32 is more valuable than item 33. The option '33 when I could have had 32' is a different option from the simple option '33', no matter how internally complex option 33 may be – and the same goes for option 32. But if my ranking order does not even commit me to claims about the relative values of the items ranked, it is pointless.

5. HOLISM AND EXPLANATION

There remains one further matter that I think it worth bringing out. It seems to me that particularism commits one to a highly debatable doctrine in the theory of explanation. In *Moral Reasons* I was not so clear about this.

I start by considering the relations between two doctrines. One is the now familiar holism in the theory of reasons. The other is a doctrine in the theory of explanation, which has no agreed name that I know of. Here they are:

> *Holism*: the ability of a consideration to stand as a reason for action can be affected by the context in which it occurs.
>
> *Non-guaranteeing explanations*: an explanation can be perfectly good without being 'complete', where a complete explanation is one that is inconsistent with the non-occurrence of the explanandum. Where E and O both occur, the occurrence of E can explain that of O without guaranteeing it. Perfect explanations can be 'non-complete'. The idea that non-complete explanations are enthymematic is a mistake.

The question that I want to start with is whether I was right or wrong to think, as I once did, that my doctrine of reasons-holism was effectively the same as this doctrine in the theory of explanation. If it is, we have uncovered another awkward consequence of particularism.

One preliminary first. Some philosophers maintain that though guaranteeing explanations are not required when what we are explaining is an action, they are required when we are explaining events. A scientific explanation, then, will have to be a guaranteeing one.[20] My own view, however, is that we would need a lot of persuasion to say any such thing. The only reason for adopting it, I think, is a sense that we would do well not to demand something of action-explanations that we have no prospect of achieving; in science, by contrast, where prospects are better, we should not be satisfied with anything less than perfection. The proper riposte to this is that action-explanations are as good as any explanation needs to be, and that, so far as the purposes of explanation are concerned, there is no reason to think there is

[20] For a recent example of this view, see Philippa Foot, 'Does Moral Subjectivism Rest on a Mistake?', *Oxford Journal of Legal Studies,* 15, 1 (1995), 1-14.

either need or room for anything better. We should therefore accept the possibility of non-guaranteeing explanations on both sides or on neither.[21]

The concept of a non-guaranteeing explanation requires that of an enabling condition. If there is a non-guaranteeing explanation F of an event E, there must be an event O such that:

1. The occurrence of O is not part of the explanation of E, and
2. If O were to fail, we would have a situation (F, not-E).

In such a situation, O would be an enabling condition for the explanation that F gives us of E. The reason why there must be such things as enabling conditions if there are to be non-guaranteeing explanations is that if every candidate for the role of enabling conditions were to turn out to be part of the explanation of the event-type E, all explanations would be guaranteeing ones (when 'complete'). If we think that there are *any* non-guaranteeing explanations, then, irrespective of whether we think that no explanations are guarantors, we must be able to make sense of the notion of an enabling condition – a condition whose satisfaction is required for the explanation, but which is not itself a part of that explanation. And this sets us something of a challenge.

But we face the same challenge in the theory of reasons for action, once we adopt holism there. This is not perhaps too surprising antecedently, since presumably the reasons favouring an action are reasons that explain that action's rightness (or whatever moral status it has). If we start from one case where there are reasons R1-Rn making the action right, but allow that changes elsewhere may affect the ability of those reasons, not merely to make the action right, but even to be reasons at all, we will again say that features over and above the reasons must be functioning as enabling conditions. They enable the features that are reasons to be the reasons they are in this case, without themselves being among the reasons why the action is right.

To give a very simple example: suppose that ought implies can. Then if I cannot do the action, the features which, were I capable of doing it, would be reasons why I should do it, are incapable of playing that role. But we should not conclude that my ability to do it is one of the reasons why I should do it. It is a condition that enables the reasons why I should to be the reasons they are, but is not itself among those reasons. Allow this, and you will probably allow the next: that I have the opportunity to help is something without which the reasons why I should help would not be those reasons. Her need is not a reason for me to help her if I have no possible opportunity of doing so. It is only a reason for me to seek an opportunity, which is different. But that I have an opportunity to help is not itself among the reasons for doing so. Another similar example: if I were not alive, the reasons that there are for me to help the needy would not be able to be the reasons they are. But this does little

[21] My views on the nature of explanation in general bear interesting similarities to those of Nancy Cartwright. See her *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983).

to show that among the reasons why I should help the needy is the fact that I am alive.

So we see the same structure both times. Both holism and the claim that explanations need not guarantee are committed to making sense of the notion of an enabling condition. One might think[22] that there are forms of holism that don't have this effect. We might try to adopt a sort of weak holism without adopting any distinctive doctrine in the theory of explanation.

> *Weak holism*: the ability of a consideration to stand as a reason can be affected by what other features are present as reasons (but not by other things).

With this in hand, we might hope to avoid the need to talk about enabling conditions at all, and avoid the need to allow non-guaranteeing explanations. But we can only do this if we can add to our weak holism a sort of:

> *Holistic generalism*: taking all the reasons here present together, they guarantee the rightness of the action; i.e. where present together elsewhere, they will always have the same effect (non-causally, of course).

This holistic doctrine is certainly compatible with the denial of a different form of generalism:

> *Atomistic generalism*: each reason has the same tendency as a reason, no matter what the context in which it is found.

But in fact weak holism cannot be coherently combined with holistic generalism. If the ability of a consideration to stand as a reason can be affected by what other features are present as reasons, why is it that the presence of further reasons in a second case, in addition to all those present in the first, is incapable of making any difference to the original reasons? Surely the official statement of weak holism says that a new reason can upset others whether we take them one by one or all together. So the combination of weak holism and holistic generalism is not a sustainable position.

Moving now to the theory of explanation, we find the same thing. There might seem to be a form of holistic generalism available there, that is compatible with the demand that all explanations be guaranteeing explanations. We might say, that is, that though no individual feature has its own explanatory potency, which it carries with it from case to case, regardless of changes in other explanatorily potent features, still the entire complex of features that together explain the event constitutes a guaranteeing explanation. It is a guaranteeing explanation because that

[22] And I did think so, until Eve Garrard showed me that I shouldn't.

complex of features could not have occurred unless the event to be explained had occurred (or been going to occur).

But this attempt to make room for guaranteeing explanations within a holistic picture suffers from the same incoherence that we found in the theory of reasons. No explanation is given of why the arrival on the scene of a new element is capable of affecting the explanatory contributions of individual features, but not of a set of such features. Why is it that a whole set of explanatory features is necessarily invulnerable to the sort of difference that a new feature can make to individual elements in that set? There is just no answer to this question. So there is no way for a holist to avoid making room for the concept of enabling conditions, and denying the possibility of guaranteeing explanations.[23]

One useful consequence of these thoughts is that the incoherence of the combination of weak holism and holistic generalism gives us an answer to one leading reply to particularism.[24] This reply is that my examples only concern what we might call *contributing* reasons. It is true that contributing reasons do vary in their polarity from one context to another. But it is not true of *complete* reasons. These remain the same, as reasons, regardless of changes elsewhere.

There are two ways in which I can respond to this. The first is, as above, to point out that there is no obvious explanation of why a new reason in a new case is supposedly able to change the behaviour of one reason but not of all of them at once. Why is it that the whole pack of them is immune to change, when no individual one is? Note that this question is asked within the constraints of weak holism; it would be a different matter to maintain that considerations that are not reasons in the new case are able to make changes in this way.

The second way in which I can respond to this attack is to challenge its notion of a complete reason. This cannot be identical with all the reasons present in the particular case, since we have already seen that there is no justification for the view that the pack of all such reasons is invulnerable to changes brought about by the presence of a new reason in a new case. It must therefore be something greater than that; it will presumably contain all the enabling conditions, as well as the absence of disabling conditions (specified one by one). But, first, this collection is ceasing to look like a reason at all, and, second, there is beginning to be a prospect that a complete reason will expand indefinitely.[25]

---

[23] The same may be true about causal statements: a sufficient cause need not be a guaranteeing cause. But to try to argue this would take us too far away from present concerns.

[24] I mentioned this reply earlier, in discussing holism in theoretical reason. A recent example is to be found in Jonathan Bennett, *The Act Itself* (Oxford: Clarendon Press, 1995), 80.

[25] Jonathan Lowe pointed out to me that a valid deductive argument is something we could reasonably call a complete reason, since it is monotonic, i.e. no addition of premises will affect the validity of the argument. In this, deductive reasoning differs from probabilistic reasoning. So there is one model of a complete reason that I cannot undermine. This model is, however, not applicable to moral reasoning, unless there are absolute moral principles; reasoning that runs in terms of prima facie principles is non-monotonic. There is certainly available a notion of deductive reasoning in ethics. The only question is whether the reasoning that takes us from premises specifying the features that make an action right to the conclusion that it is right is ever deductively valid. (I mean by this form of words to exclude such

What I really want to say, of course, is just that my interests are in the ordinary notion of a contributing reason, not in the concocted notion of a complete reason, which in my view is designed merely to save a dubious philosophical theory from refutation. But this may seem to be a definitional sulk; so I support it by attacking the notion of a complete reason that is designed to save generalism from counter-example.

premises as 'all actions of this sort are right', 'this feature is the only relevant one' and 'this feature is a *pro tanto* reason'.) I think not.

CHAPTER 28

G. H. HARMAN

TOWARD A THEORY OF INTRINSIC VALUE[*]

Any complete normative theory of value will include an account of when something's being the case would or would not be a "good thing," a "worth-while thing," or a "desirable thing." Some philosophers believe that in order to give such an account one need (and need only) say when something's being the case would or would not be *intrinsically* good, worth while, or desirable. To do this would be to present a theory of *intrinsic value.* In this paper I examine what I will call "the standard account" of intrinsic value as it appears in recent textbooks written by John Hospers, William Frankena, and Richard B. Brandt.[1] I argue: (a) it is not clear whether a theory of intrinsic value can be developed along the lines of the standard account; (b) if one is to develop such a theory, one will need to introduce a notion of "basic intrinsic value" in addition to the notion of "intrinsic value"; and (c) several different theories of intrinsic value may account for the same judgments concerning desirability, and it will be arbitrary to choose one of these theories over another.

1. PRELIMINARY REMARKS

For the purposes of this paper I consider the following expressions to be equivalent: "*X* is (or would be) a good thing," "*X* is (or would be) a worth-while thing," "*X* is (or would be) a desirable thing," "*X* is (or would be) good," "*X* is (or would be) desirable," "*X* is (or would be) worth while," "*X* has (or would have) intrinsic value." I am interested only in cases in which the '*X*' of these sentence forms is replaced with the nominalized form of some sentence. For example, taking "The cat is on the mat" as the sentence to be nominalized, the first of these sentence forms could be realized in any of the following ways: "That the cat is on the mat is a good thing," "That the cat would be on the mat would be a good thing," "For the cat to be

---

[*] Some of the ideas presented here were suggested to me by Robert Nozick. I have also benefited from suggestions by Richard Jeffrey and Monroe Beardsley, who read an earlier draft. Responsibility for errors is my own.

[1] John Hospers, *Human Conduct* (New York: Harcourt, Brace, & World, 1961), pp. 104-105; William Frankena, *Ethics* (Englewood Cliffs. N.J.: Prentice-Hall, 1963), p. 67; Richard B. Brandt, *Value and Obligation* (New York: Harcourt, Brace, & World, 1963), pp. 18-20; Brandt, *Ethical Theory* (Englewood Cliffs, N.J.: Prentice-Hall, 1959), p. 302. Parenthetical page references are to these books.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 349-360.
© 2005 *Springer. Printed in the Netherlands.*

on the mat is a good thing," "The cat's being on the mat is a good thing," etc. Again for the purposes of this paper I consider all these sentences to be equivalent.[2]

Thus, I am not concerned with cases in which the '*X*' is replaced by something that refers to an object in a restricted sense of 'object'. For example, in a recent article on intrinsic value, Monroe Beardsley discusses the value of a rare postage stamp.[3] I do not think that the usual philosophical notion of intrinsic value can have any application to such a case. It makes sense to ask whether my having such a stamp would be intrinsically worth while or whether such a stamp's existing would be intrinsically desirable. But, since it does not seem to make sense to ask whether such a stamp is itself intrinsically valuable, I will not be interested in discussing the value of objects like stamps, per se.

I provisionally accept a distinction between the normative theory of value and a metaethical theory of value. I understand a normative theory of value to be a systematic account of when something's being the case would be a good thing. A metaethical theory is a theory about the meaning of value judgments in ordinary language and, possibly, a theory about the way in which these judgments are to be "proved" or "validated." I take the notion of intrinsic value to be a technical notion introduced to simplify the normative theory of value: once one has specified what has intrinsic value, one can then say that anything has value to the extent that it "leads to" something having intrinsic value. I doubt that the notion of intrinsic value plays an important role in metaethics. Thus, I do not think that it is intuitively obvious what things have intrinsic value; nor do I believe one ordinarily decides whether something's being the case would have value by discovering whether it would "lead to" intrinsically desirable things.

William Frankena disagrees. In the following passage he says we ordinarily determine what things have value on the basis of what things have intrinsic value:

> But it still remains true that in order to come to a judgment about whether something is good on the whole or good in any of the other senses, we must first determine what its intrinsic value is, what the intrinsic value of its consequences or of the experience of contemplating it is, or how much it contributes to the intrinsically good life (67).

This very strong claim about intrinsic value is refuted by Beardsley's observation: "The paradoxical feature of our value knowledge is just that we have a good deal of sound knowledge about instrumental values, but are in considerable doubt about intrinsic values" (7). Richard Brandt is rightly more cautious than Frankena:

> It would seem, however, that if anything is desirable at all, it must be because some facts or events (perhaps the thing itself) are intrinsically desirable. But a person would not be contradicting himself if he denied this statement, and it is

---

[2] I am not concerned with what such nominalized sentences "refer to," e.g., states of affairs, possible events, sentences, or sets of sentences.

[3] Monroe C. Beardsley, "Intrinsic Value," *Philosophy and Phenomenological Research,* XXVI (September 1965): 1-17 [* pp. 61-75 of this volume]. Parenthetical references to Beardsley are to this article. Reference here: 1-2 [* pp. 61-62 of this volume].

> an interesting question for reflection what one should say to a person who did deny it (*Ethical Theory*, 302).

In this paper I attempt to explain why it can seem that what is desirable depends upon what is intrinsically desirable; and I try to specify what would be involved in denying this claim.

2. THE STANDARD ACCOUNT

John Hospers introduces the notion of intrinsic value in a standard way:

> Our question, "What things are good?" is ambiguous. There are some things which we consider good (or desirable) only for their results – for what they lead to. There are other things which we consider good not because of what they lead to but because of what they are in themselves: we consider them worth having or pursuing not merely as ways of getting other things but because of their own intrinsic nature. The first kind of good is called *instrumental* good because the goodness or worth-whileness of these things lies in their being instruments toward the attainment of the other things which are considered good not merely as instruments. The second kind of good is called *intrinsic* good because we value these things (whatever they may turn out to be) not for what they lead to but for what they are (104-105).

Brandt, in *Ethical Theory*, says, "To say that something is intrinsically desirable is to say that it is desirable, taken just for itself, viewed abstractly, and in particular, viewed without respect to any consequences its existence will or may produce" (302). In *Value and Obligation*, where Brandt offers what he calls "a rough definition of 'intrinsic value'," he says, "Something is intrinsically desirable if it would be properly valued or desired or chosen for its own sake, when considered independently of its consequences" (18). His more precise explanation says:

> Some facts or events, then are desirable for no reason beyond their being what they are; they are solely intrinsically desirable. Other facts or events are desirable solely for reasons beyond themselves, usually their effects; they are not intrinsically desirable at all. Still other things are desirable partly just because of what they are but partly because of other things, preeminently their effects; they are in part intrinsically desirable but in part desirable for other reasons (19).

Each account obscurely distinguishes two sorts of reasons for which something may be desirable, Hospers says the one sort of reason mentions what things "lead to"; the other sort mentions "what they are." But how are we to tell when a reason mentions what *S* is rather than what *S* leads to? Similarly, how are we to understand Brandt's distinction between reasons for things' being desirable which are "their being what they are" and reasons for things' being desirable which are "reasons beyond themselves"?

Brandt answers these questions by saying, of the "facts" that might be cited as reasons:

> Now such facts that make an event desirable may or may not be different from the very fact or event originally said to be desirable and may be different from anything included in, or logically required or necessitated by, this fact or event (*Value and Obligation*, 19).

I think Brandt's suggestion can be interpreted as follows: a reason (for *S*'s being desirable) refers to "what *S* is" and is, therefore, "included in *S*," if and only if the reason is that something is logically required or necessitated by *S*; so *S* is intrinsically desirable to the extent that what makes *S* desirable is that *S* entails something, *T*. *S* fails to be intrinsically desirable and is instrumentally desirable to the extent that what makes *S* desirable is something other than *S*'s entailing something, (This interpretation does not quite fit what Brandt says. Instead of saying that *S* is intrinsically desirable if what makes *S* desirable is that *S* entails something, Brandt seems to say that *S* is intrinsically desirable if *S* entails what makes *S* desirable. However, this last remark is obscure when taken literally: and I think Brandt meant to say what I interpret him as saying.)

It still remains to be explained how a reason why *S* is desirable can be *what makes S desirable* in the required sense. There are reasons why *S* is desirable that are not the sort of reason Brandt must have in mind. For example, if something is a reason why *S* is desirable in that it is the causal reason why *S* will have desirable consequences this time, then this reason would not be what Brandt has in mind when he speaks of *what makes S desirable.* Again, if something is a reason why *S* is desirable in that it provides a reason for inferring that *S* is desirable (e.g., the reason is that someone reliable has said that *S* is desirable), then such a reason would also not be what Brandt has in mind. Brandt suggests something of the nature of the relevant sort of reason when he equates 'desirable' with 'properly desired or chosen' and remarks: "Something is intrinsically desirable if it would be properly desired or chosen for its own sake" (18). This indicates that the relevant sort of reason must be a *reason for desiring or choosing* the thing in question.

If *S* is not already desired, reasons for desiring *S* typically cite *S*'s connection with something that is already desired. The connection may be that *S* entails something already desired or that *S* would "lead to" something already desired. We want to adapt these remarks to desirability; although we cannot of course make use of a literal distinction between what is already desirable and what is not already desirable (to parallel the distinction between what is already desired and what is not already desired). In effect, the standard account identifies what has intrinsic value with what is already properly desired. This works well for instrumental desirability: if *S* does not entail but "leads to" something intrinsically desirable, then *S* is instrumentally desirable. However, if *S* entails something having intrinsic value, we want to be able to say that *S* is itself intrinsically desirable. Perhaps the best way to handle this case is to say that, if *S* entails something already properly desired, then *S* is already implicitly properly desired and is therefore intrinsically desirable. (The awkwardness of this maneuver arises from the fact that the real parallel is not between what is already desired and what is intrinsically desirable but is rather between what is already desired and what is basically intrinsically desirable. The

notion of basic intrinsic value is discussed in section V, below.) The standard account says:

(1) *S* is intrinsically desirable to the extent that *S* entails something that is intrinsically desirable
(2) *S* is instrumentally desirable to the extent that *S* does not entail but "leads to" something intrinsically desirable.

Notice that neither of these propositions permits a noncircular *definition* of 'intrinsic value' or of 'instrumental value'.

3. AN IMPROVEMENT IN THE STANDARD ACCOUNT

Hospers' introduction of the term 'intrinsic value', quoted above, suggests that the second type of reason for properly desiring *S* must be that *S* would "lead to" things of intrinsic value. It is natural to interpret "*S* would lead to *T*" to mean "*S* would cause or result in *T*" or "*T* would be an effect of *S*." Brandt is nearer the truth. He says the reasons for which things are instrumentally desirable are only "preeminently their effects" (19). (I understand him to mean that the reasons are preeminently that these things have certain effects.) Even Brandt's claim is too strong. An important way in which *T* may provide a reason for properly desiring *S* is that *S* increases the likelihood of *T*, where *T* is desirable. *S* may be a *sign* of something desirable or of something undesirable, without necessarily being a cause of something desirable or undesirable. For example, whether or not it would be a good thing for me to receive one hundred thousand dollars sometime this year depends heavily on how it happens that I receive this money. If the only probable way in which this could happen requires the death of some member of my family so that I receive insurance, then I cannot think of receiving this money as a desirable thing.[4] So, the desirability of *S* will be affected by the desirability of those things which *S* makes more or less probable. Furthermore, just how the desirability of *S* will be affected depends upon just how desirable these things are and just how *S* would affect their likelihood. To spell this out will require our being able to speak meaningfully of assigning some measure to desirability.

Even if we were to restrict our attention to *S*'s effects, it would be an oversimplification to say that *S* is desirable because of what it would lead to. Instead we must say that *S* is desirable because of what it is *likely* to lead to. *S* may happen in various ways having different consequences. For example, your giving me ten dollars will have varying consequences depending on whether you give me the money as a gift, to hold for the purposes of a bet, or to give to someone else. Its consequences will also depend upon whether you give me the money in pennies, in dollar bills, or in the form of a check (Jeffrey 190-192, 196-199). *S*'s being desirable

[4] Cf. Richard Jeffrey, *The Logic of Decision* (New York: McGraw-Hill, 1965), p. 198. Parenthetical page references to Jeffrey are to this book.

because of what it would lead to must be taken as a special case of *S*'s being desirable because of what it would make more likely.

Judgments about what has already happened are only apparent counterinstances to this claim. It is true that, when we judge Caesar's crossing the Rubicon to be a good thing, we base this judgment on the actual consequences of his crossing the Rubicon and not on the probable consequences of this crossing. It is also true that, relative to our present knowledge of those consequences, we cannot say that Caesar's crossing made or would make these consequences probable. However, the judgment that Caesar's crossing the Rubicon was a good thing amounts to the judgment that it was better than his not crossing the Rubicon. To evaluate this latter claim involves an estimate of what the consequences of his not crossing would have been; and, since there are various ways in which it might have happened that he did not cross, one would have to estimate the probabilities of various consequences in a way similar to that one uses to estimate the probabilities of consequences of something that has not yet happened. Although past events are not judged desirable or undesirable because of what consequences they made likely, they are judged desirable or undesirable because of what consequences alternatives would have made more or less likely. In what follows I will not make use of the special formulations needed because of the indirect and counterfactual character of judgments about the past.

Therefore, the improved standard account of intrinsic value says:

(1) *S* is intrinsically desirable to the extent that *S* entails something that is intrinsically desirable.

(2') *S* is instrumentally desirable to the extent that *S* does not entail but increases the probability of something intrinsically desirable.

4. AN UNEXPECTED PROBLEM

Nevertheless, we are closer to being able to say just what is involved in a theory of intrinsic value. According to this theory, the value or desirability of something *S* depends upon *S*'s intrinsic value and also upon the intrinsic value of those things *S* makes more or less probable. Given (1) and (2'), this amounts to the claim that all value is either intrinsic or instrumental or both. In order to specify the theory further we must be precise about what contribution something else can make to the value of *S*. Suppose that we are able to make sense of the notion of a definite *amount* of value. Then it would seem natural to say that *T*'s contribution to the value of *S* is the intrinsic value of *T* multiplied by the amount that *S* increases the probability of *T*. But this will not do.

Let V(*X*) be the over-all value of anything *X*. Let I(*X*) be the intrinsic value of *X*. Let PM(*X*/*Y*) be the amount that *Y* increases the probability of *X*.[5] For any function *F*

[5] If P(*X/Y*) is the conditional probability of *X*, given *Y*, then PM(*X/Y*) is P(*X/Y*) minus P(*X*/(*Y* or not-*Y*)).

whose values are real numbers, positive or negative, let $\Sigma_X F(X)$ be the sum of the values of F for every possible value of the argument X. Suppose also that S and T are identical if and only if they entail each other. Then the proposal under discussion is

(3) For any S, $V(S) = \Sigma_T [I(T) \times PM(T/S)]$

The difficulty with the proposal is that, according to (1), something may have intrinsic value because it entails something else of intrinsic value. For example, suppose that *S* has intrinsic value. Then so does *S&T*, since *S&T* entails *S*. But, once we have given *S&T* credit for increasing the probability of something that has intrinsic value in its own right, we do not want to give *S&T* further credit for increasing the probability of something else, itself, that has intrinsic value just because it entails the first thing. Thus, suppose that someone's getting pleasure at noon today would have an intrinsic value of 10 units and has an initial probability of .50. Then my getting pleasure at noon today would also have an intrinsic value of 10 units, and an initial probability of .50 or less, because it entails someone's getting pleasure. And my getting pleasure at noon today in this room would also have an intrinsic value of 10 units, and an initial probability of .50 or less, for the same reason. My getting pleasure at noon in this room thus increases the probability by at least .50 of at least three things, each possessing 10 units of intrinsic value: itself, my getting pleasure at noon, and someone's getting pleasure at noon. Using (3), my getting pleasure at noon would have to be worth at least 15 units of value. But this result must be wrong.

5. BASIC INTRINSIC VALUE

The example suggests that something may have intrinsic value in a more or less basic way. Someone's getting pleasure at noon today might have intrinsic value in the most basic way, whereas my getting pleasure at noon today in this room may have intrinsic value only in a derivative way, because it entails something that has intrinsic value in the most basic way. Something made more or less probable by *S* contributes to the value of *S* only if it possesses intrinsic value in the most basic way. If *S* makes probable my getting pleasure at noon today in this room and if this has intrinsic value only because it entails someone's getting pleasure at noon today, then we want to give *S* credit only for increasing the probability of someone's getting pleasure at noon today. It would be an error to give *S* further credit for increasing the probability of my getting pleasure at noon today in this room.

Therefore the theory of intrinsic value requires the introduction of the notion of *basic intrinsic value* and a basic intrinsic value function B satisfying two requirements. Let $\Sigma_{S,T} F(T)$ be the sum of the values of F for all arguments T such that S entails T. Then the intrinsic value of S is given by

(4) $\mathrm{I}(S) = \Sigma_{S,T}\mathrm{B}(T)$

The total value of *S* is given by

(5) $\mathrm{V}(S) = \Sigma_{T}[\mathrm{B}(T) \times \mathrm{PM}(T/S)]$

So, the following statement represents a necessary condition of the truth of any theory of intrinsic value.

(6) There is a function B, the *basic-intrinsic-value function*, such that the intrinsic value of anything *S* is given by (4) and the value of *S* is given by (5).

(6) is a strong thesis. First, (6) presupposes that there is a measure of desirability susceptible to the sort of addition and multiplication envisioned in (4) and (5). Second, it claims that there is at least one function, B, determining the values of everything by satisfying (5). Third, it claims that, of the functions satisfying (5), at least one represents basic *intrinsic* value in that it satisfies (4). None of these claims is obviously true. So it cannot be obvious that anything has intrinsic value.

6. A FIRST-CAUSE ARGUMENT

There is a familiar argument for the claim that some things must be intrinsically desirable. (Beardsley (6) calls this argument, "The Dialectical Demonstration.") It is as follows:

> Consider something *A* that is desirable. Either *A* is desirable for itself, and therefore something is intrinsically desirable, or *A* is desirable only because *A* leads to something else, *B*, that is desirable. Then either *B* is intrinsically desirable or *B* is desirable only because *B* leads to something else, *C*, that is desirable. And so on. But if nothing is intrinsically desirable, how can *A* ever be desirable, since *A* is desirable only because *B* is desirable, *B* is desirable only because *C* is desirable, and so on *ad infinitum,* with no explanation why *any* of the members of this series is desirable? (6 [* pp. 65-66 of this volume])

As it stands, this argument is no more compelling than the first-cause argument for the existence of God. There is nothing inconsistent in the notion of an infinite series *A*, *B*, *C*, etc., each member of which is desirable only because the next is. However, it is by now easy to see why one can think it obvious that something must have intrinsic value and why one might be tempted to argue for this using the "first-cause argument" just outlined, if one identifies intrinsic value with basic intrinsic value. Suppose there is a basic intrinsic value function B satisfying (5). If nothing had basic intrinsic value, then $\mathrm{B}(T) = 0$, for all *T*, and $\mathrm{V}(S) = \Sigma_{T}[0 \times \mathrm{PM}(T/S)] = 0$, for arbitrary *S*. So, if nothing had basic intrinsic value, then (as the proponents of the

"first-cause" argument maintain) nothing would have any value at all. Since something does have value, it would follow that something must have basic intrinsic value. Of course, this conclusion follows only because one has assumed that there exists a basic-intrinsic-value function B, satisfying (5). Since this assumption is not obviously true, it is not obviously true that something has basic intrinsic value.

7. AN ARGUMENT AGAINST THE THEORY OF INTRINSIC VALUE

Consideration of the following arguments helps to clarify what is involved in the theory of intrinsic value. The first is a familiar argument against the theory that all pleasure and only pleasure is of intrinsic value:

> Suppose that during a war a prisoner, Smith, is being tortured in an attempt to extract information from him. His torturer, Jones, is someone who has never done this sort of thing before. Both Jones and Smith are about to die when a bomb destroys their building after a direct hit. Call these circumstances *C*. Then compare the relative desirabilities of *S* and *S′*, where *S* is that in *C* Jones enjoys himself, and where *S′* is that in *C* Jones does not enjoy himself. *S* and *S′* are quite similar in what they make more or less probable, except that *S* entails that Jones will have pleasure that *S′* entails Jones will not have. On the assumption that all and only pleasure is of intrinsic worth, *S* ought to be more desirable than *S′*. But *S* is less desirable than *S′*. (Suppose for the sake of argument that this premise is true.) Therefore, the assumption that all and only pleasure is of intrinsic worth must be false.

A similar argument against the possibility of any basic intrinsic value function B continues as follows:

> Suppose there is a function B satisfying (5). Then,
>
> $$V(S) = \Sigma_T[B(T) \times PM(T/S)] \text{ and } V(S') = \Sigma_T[B(T) \times PM(T/S')]$$
>
> The difference between these sums must result from the difference between *S* and *S′*. Since this difference is solely that *S* entails Jones will have pleasure that *S′* entails Jones will not have, it follows that the respective sums will be exactly the same, except that the first sum will contain terms representing the value of Jones's pleasure, whereas the second sum will not contain these terms. Thus, $V(S) = \Sigma_T[B(T) \times PM(T/S)] = \Sigma_T[B(T) \times PM(T/S')]$ + (terms representing the value of Jones's pleasure) = $V(S')$ + (terms representing the value of Jones's pleasure). The terms representing the value of Jones's pleasure are those terms in the sum for the value of *P*, where *P* is that Jones enjoys

> himself during a certain interval of time; so V(S) = V(S') + V(P). Since V(P) > 0, it follows that V(S) > V(S'). But, by hypothesis, V(S) < V(S'). The original assumption that there is a basic-intrinsic-value function B has led us into contradiction; so that assumption must be incorrect.

The first argument, against the view that all and only pleasure is intrinsically good, seems valid. (The truth of the premise that S is less desirable than S' is another matter.) The second argument, against the view that there is a basic-intrinsic-value function B, goes wrong in claiming that the terms representing the value of Jones's pleasure are simply the terms in the sum for V(P). In fact, there need be no simple relationship between the sums for V(S) and V(S'). The slight difference between S and S' may make a great deal of difference in what S and S' make more or less probable. S' makes it certain that Jones is not taking pleasure in Smith's suffering, and P (i.e., Jones's getting pleasure over a certain interval of time) does not, taken by itself, make it highly probable that Jones is taking pleasure in Smith's suffering; yet S does make it highly probable that Jones is taking pleasure in Smith's suffering. Although S makes it certain that Jones is getting pleasure, something to which the basic-intrinsic-value function will perhaps assign a positive value, S also makes it highly probable that Jones is getting sadistic pleasure, and the basic-intrinsic-value function might assign this a low enough intrinsic value so that, as a result, S ends up with a lower total value than S'. Therefore, the above argument cannot be used to show that there is no basic intrinsic value function B.

8. POSSIBLE ARBITRARINESS IN THE THEORY OF INTRINSIC VALUE

Beardsley offers a different type of argument against the theory of intrinsic value. He claims that "the concept of intrinsic value is inapplicable – that, even if something has intrinsic value, we could not know it, and therefore that it can play no role in ethical or aesthetic reasoning" (12-13 [* p. 71 of this volume]). His argument is that, since we cannot discover the intrinsic desirability of S by discovering that S would lead to other desirable things, the only other way to discover S's intrinsic desirability would be via immediate intuition. He takes this consequence to discredit the theory of intrinsic value.

But it is not true that the only way to show that something is desirable is to show that it would lead to (i.e., make more or less probable) something else. Another way to show that something is desirable is to show how the assumption that it is desirable would explain why other things are desirable. In other words, we might use in ethics the same sort of argument we would use elsewhere, an argument that something is the best explanation of our data.[6] And that is just the way in which one would want to defend a theory of intrinsic value. One would want to argue that it is possible to

[6] See the author's "The Inference to the Best Explanation," *The Philosophical Review,* LXXIV, 1 (1965), pp. 88-95.

simplify and thus explain the theory of value by assuming, for example, that someone's getting pleasure is always of basic intrinsic value. Beardsley has not shown that such an argument could not be given, for some theory of intrinsic value. Therefore, he has not shown that "even if something has intrinsic value, we could not know it."

Nevertheless, it is not obvious that the theory of intrinsic value can be defended in the way just suggested. Beardsley points out (in personal communication) that there may be more than one function B, satisfying (5). If so, then these different B's may define different intrinsic-value functions I, by virtue of (4), and one would need to have some way of deciding which such function was the *real* intrinsic-value function. (Of course, there may be several functions B, satisfying (5), each defining what turns out to be the same intrinsic-value function by virtue of (4).) For example, suppose one wants to defend the view that getting pleasure is always and only of intrinsic value. It would seem that the same values of V will be determined if we take someone's getting pleasure to be of basic intrinsic value, or if we take it, that for any person *P*, *P*'s getting pleasure has basic intrinsic value. This would seem to be a case where different B's might satisfy (5). But (4) will define different intrinsic-value functions depending on which B is selected. If we say that, for any *P*, *P*'s getting pleasure is of basic intrinsic value and is the only sort of thing that is, then it will not follow that *either Jack's getting pleasure or Bill's getting pleasure* will be of intrinsic value, since, for any *P* this does not entail *P*'s getting pleasure. On the other hand, if we say that someone's getting pleasure is of basic intrinsic value, then *either Jack's getting pleasure or Bill's getting pleasure* will be of intrinsic value, since it entails someone's getting pleasure. Therefore, there may be several intrinsic-value functions satisfying (4) and (5). But if there are several such functions, nothing said so far can choose among them, since all the restrictions stated so far have been encapsulated in (4) and (5).

This suggests that we ought to give up the idea of a unique intrinsic-value function giving the *real* intrinsic value of various things. Instead we ought to take the choice among the functions satisfying (4) and (5) to be arbitrary in the way that the choice of axioms for a formalized theory is arbitrary. A particular choice of axioms is made for reasons of simplicity and economy; and one would want to choose the basic intrinsic-value function for similar reasons.

One might try to reduce arbitrariness by making use of Brandt's equation: desirable = properly desired. But there is no reason to think one will succeed. For example, one might distinguish intrinsic desires from instrumental desires, and equate intrinsically desirable with properly desired intrinsically. There seem to be two ways, neither convincing, in which one might attempt to explain the notion of an intrinsic desire. First one might suggest, as Brandt does in *Hopi Ethics*,[7] that there are distinct "intrinsic attitudes" and that intrinsic desire is one of them. However, there do not seem to be such attitudes. Second, one might attempt to explain all of a particular person's desires in terms of certain "basic desires." But there is no reason

---

[7] Chicago: The University Press, 1954; pp. 164-165.

to think that a unique set of such basic desires accounts for the other desires a person has. Taking a particular set of desires to be basic may be as arbitrary as choice of a basic intrinsic-value function.

9. CONCLUSION

There are, no doubt, other problems with the conception of intrinsic value outlined in this paper. For example, if an infinite number of things have basic intrinsic value, then the sum in (5) may not be well defined. This problem would arise if, following a suggestion of Richard Jeffrey's (196-199), we took the things having basic intrinsic value to be represented by what Jeffrey refers to as "novels." These are infinite sets of sentences such that, for every *S*, either *S* or not-*S* is included in any particular set. Since, in general, *S* will increase the probability of any novel by an infinitesimal amount, the sum in (5) cannot be used to give V(*S*) on this proposal. On the other hand, there is no reason why we must accept this proposal.

I cannot think of a way to show there is a function B, satisfying (5); nor can I think of a way to show there is no such function. Therefore, it seems to me that the next thing that needs to be done to develop an adequate theory of intrinsic value is to discover such a function B, or prove that that there can be no such function. Until this has been done, one cannot evaluate the claim made by Frankena, Brandt, and Hospers, and possibly denied by Beardsley, that the notion of intrinsic value enables us to simplify the theory of value.

CHAPTER 29

E. CARLSON

THE INTRINSIC VALUE OF NON-BASIC STATES OF AFFAIRS

Even if we could agree on an axiological theory which singles out certain states of affairs as 'evaluatively basic' and assigns intrinsic values to those basic states, there would still be a problem about the assignment of intrinsic value to states that are not basic. Here I shall first criticize an approach to this problem developed by Warren Quinn and Edward Oldfield, and then state and discuss three alternative solutions of my own.

1.

I shall assume that intrinsic value is possessed by states of affairs, and that states of affairs resemble propositions, at least in the respect that truth-functional connectives can be meaningfully applied to them. Thus, it makes sense to speak of, for example, conjunctive, disjunctive, and negative states. As a criterion of identity, I shall suppose that two states are identical if and only if (iff) they are logically equivalent.[1] Further, let us presuppose the concept of an 'evaluatively basic' state of affairs. The intrinsic values of non-basic states are determined by the intrinsic values of the basic states. A basic state is such that its intrinsic value can be known without knowing the intrinsic value of any other state.

According to Warren Quinn, the most likely candidates for being basic states are states "which locate a specific sentient individual at a specific point along an

[1] These assumptions seem to be shared by most philosophers who have written about the intrinsic value of non-basic states. (See e.g. Chisholm (1975), Oldfield (1977), Quinn (1974), and Sobel (1970). Quinn makes it clear that he uses the terms 'state of affairs' and 'proposition' interchangeably, and the same seems to be true of Oldfield, although he predominantly speaks about 'propositions'.) But the assumptions are of course not uncontroversial. On some theories of states of affairs, there may not be such things as disjunctive and negative states. And even if there are such states, it might be denied that logical equivalence is the correct criterion of identity. (For a discussion of some different criteria of identity for states of affairs, see Åqvist (1968), pp. 259-60 and 265-7.)

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 361-370.
© 2005 *Springer. Printed in the Netherlands.*

evaluatively relevant dimension such as happiness, virtue, wisdom, etc." (Quinn (1974), p. 131.) I shall adopt Quinn's idea, with one qualification. A basic state should, I believe, locate an individual at a point on an evaluatively relevant dimension *during a certain time-interval*. Lacking such a time-reference a state such as 'Jane is happy to degree 2' does not seem to have a determinate value, and hence not to be basic, since it does not tell us for how long Jane is happy. Presumably, it is intrinsically better that she is happy for a year than that she is happy for a day. We may assume, however, that all the examples to follow concern one and the same time-interval.[2] Hence, the time-parameter need not be written out.[3]

Let us for the sake of the discussion assume that happiness is measurable (on a ratio scale) and evaluatively relevant. A basic state containing a greater quantity of happiness, or a smaller quantity of unhappiness, is thus intrinsically better than a basic state containing a smaller quantity of happiness, or a greater quantity of unhappiness. (No basic state contains both happiness and unhappiness.) That Jane is happy_2 (i.e., happy to degree 2 during the relevant time-interval) and that Bill is happy_{-4} (i.e., unhappy to degree 4 during the relevant time-interval) are examples of basic states, according to this theory. These states can be represented by 'j_2' and 'b_{-4}', respectively. Let us also suppose that for each basic state, its intrinsic value corresponds to the degree of happiness or unhappiness it involves, so that, letting 'IV' stand for 'intrinsic value', $\text{IV}(j_2)=2$, $\text{IV}(b_{-4})= -4$, and so on.

I shall make no distinction between a state's having zero intrinsic value and its lacking intrinsic value altogether. Some would perhaps want to make such a distinction, and claim that states like j_0 and j_2 & b_{-2} have zero intrinsic value, while states like 'Grass is green' and 'There are stones' lack intrinsic value. In the present connection, however, there seems to be no harm in stipulating that also the latter states have zero intrinsic value.

Assuming this simple model, our problem is how to evaluate states which are not basic. What are the intrinsic values of states such as, for example, $\neg j_2$, 'Somebody is happy_2', j_2 & b_{-4}, or $j_2 \vee b_5$? Quinn and Oldfield have suggested interesting and closely related theories about the intrinsic values of such non-basic states. A questionable feature of their approach, however, is that it makes the intrinsic values of certain states *world-relative*, i.e., different in different possible worlds. This idea of world-relative *intrinsic* value seems dubious. It is often said that the intrinsic value of a state depends wholly on the intrinsic properties of that state, or that the intrinsic value of a state is the value it has necessarily, i.e., in every possible world.

---

[2] Note that this assumption implies that, in the examples to follow, different basic states involving the same person are mutually incompatible. A person cannot be happy or unhappy to different degrees during (the whole of) one and the same time-interval.

[3] Many people would perhaps reject Quinn's presupposition that the entities located along an evaluatively relevant dimension are always 'sentient individuals'. For example, if beauty is a relevant dimension, works of art could be among the things located by basic states. And if equality is relevant, the things located might include societies or populations. Although my examples will all be about people's happiness and unhappiness, the supposition that the basic states involve sentient individuals is not essential to the discussion.

Some such requirement seems necessary to capture our intuitive distinction between intrinsic and extrinsic or instrumental value. A theory that allows world-relative intrinsic value seems in great danger of obliterating this distinction.

Apart from this general complaint, there are also more specific difficulties with Quinn's and Oldfield's proposals. The latter theory is worked out in criticism of the former and is, I believe, an improvement upon it. I shall therefore explicitly discuss only Oldfield's theory, but most of the critical remarks I am going to make apply also to Quinn's theory.

Before giving his principle for determining intrinsic value, Oldfield makes two preliminary definitions. First, a state p is *irrelevant* iff p is logically independent of every basic state, of every negation of a basic state, and of every contingent conjunctive state, each conjunct of which is either a basic state, or the negation of a basic state.[4] Second, S is a *minimal set for p in world w* iff

(i) S is a set of basic states each of which obtains in w,

(ii) there is an irrelevant state c which obtains in w and which, in conjunction with the members of S, entails p,

(iii) there is no proper subset of S which, in conjunction with some irrelevant state that obtains in w, entails p.[5]

Now we may state Oldfield's principle for the assignment of intrinsic value:

IVO For any world w and any state p,
(a) if p obtains in w, the intrinsic value of p in w is equal to the sum of the intrinsic values of the members of the union of all minimal sets for p in w,[6]
(b) if p does not obtain in w, then (i) if its intrinsic value is greater than or equal to (less than) zero in every world in which it obtains, its intrinsic value in w is equal to the least (greatest) intrinsic value it has in any world in which it obtains, and (ii) if its intrinsic value is greater than or equal to zero in some worlds in which it obtains and less than zero in some worlds in which it obtains, its intrinsic value in w is zero.[7]

Unfortunately, IVO has some rather counter-intuitive consequences. Suppose that j_1 and b_1 obtain in w. Letting 'IV_w' stand for 'intrinsic value in world w', we

[4] Oldfield (1977), p. 237. I have slightly simplified his definition. The notion of an irrelevant state is needed in order to confer intrinsic value to states like the disjunction (j_1 & 'Grass is green') $\vee$ (b_2 & 'There are stones').
[5] Oldfield (1977), p. 239.
[6] Oldfield (1977), p. 240.
[7] Oldfield (1977), p. 242.

then have that $IV_w(j_1 \vee b_1)=2$, while $IV_w(j_1 \vee b_{100})=1$.[8] But surely the latter state cannot be intrinsically *worse* than the former, since they share one disjunct and the remaining one is intrinsically *better* in the latter state than in the former. Similarly, if j_0 (i.e., the state that Jane is neither happy nor unhappy) and b_{-1} obtain in w, $IV_w(j_0 \vee b_{-100})=0$, while $IV_w(j_{100} \vee b_{-1})= -1$.[9] The latter state differs from the former in that each disjunct is replaced by a much *better* basic state. Therefore it is hardly reasonable to claim that the latter disjunction is intrinsically *worse* than the former.

IVO also yields rather peculiar results concerning universal and existential generalisations, as well as negations. Suppose that there are only ten people in w, and that they are all happy$_2$. Then IV_w('Everybody is happy$_2$')=20, while IV_w('Everybody is happy$_{19}$')=19.[10] But it is hard to believe that the former state is intrinsically better than the latter. Even more counter-intuitively, IV_w('Nobody is happy$_{19}$')=20, while IV_w('Somebody is happy$_{19}$')=19.[11] With respect to negations, IVO entails that if j_5 obtains in w, then $IV_w(\neg j_n)=5$ for all $n \neq 1$. To me it seems odd, however, that negative states like $\neg j_2$ should have a determinate (non-zero) intrinsic value at all.

Oldfield discusses the disjunctions $j_1 \vee b_{-20}$ and $j_{-1} \vee b_{20}$, and claims that the latter state is 'in a sense' intrinsically better than the former. "That sense is that even though there are worlds w_i, w_j such that the intrinsic value of $[j_1 \vee b_{-20}]$ at w_i is higher than the intrinsic value of $[j_{-1} \vee b_{20}]$ at w_j, there are also worlds at which $[j_{-1} \vee b_{20}]$ is better than $[j_1 \vee b_{-20}]$ is at any world." (Oldfield (1977), p. 245.) Let 'IV*' denote intrinsic value in this sense. I have criticized IVO for entailing, e.g., that $IV_w(j_1 \vee b_1) > IV_w(j_1 \vee b_{100})$, if j_1 and b_1 obtain in w. Oldfield might respond by pointing out that $IV^*(j_1 \vee b_{100}) > IV^*(j_1 \vee b_1)$, and that IVO thus can capture our intuition that the former state is intrinsically better, or at least not worse, than the latter.

This response is of doubtful merit. The claim that there is one sense of 'intrinsically better', according to which $j_1 \vee b_{100}$ is in w intrinsically better than $j_1 \vee b_1$, and another sense according to which the contrary is true, seems to be an *ad hoc* measure to bring his theory more in line with our intuitive judgements. Moreover, IV* is hardly a plausible notion of intrinsic betterness. For example, $IV^*(j_1 \vee j_{-100}) > IV^*(j_0)$, and IV*('Nobody is happy$_{20}$') > IV*('A million people are happy$_{20}$').[12]

---

[8] Both $\{j_1\}$ and $\{b_1\}$ are minimal sets for $j_1 \vee b_1$ in w. On the other hand, $\{j_1\}$ is the only minimal set for $j_1 \vee b_{100}$ in w, since b_{100} does not obtain in w.

[9] $\{j_0\}$ is the only minimal set for $j_0 \vee b_{-100}$ in w, while $\{b_{-1}\}$ is the only minimal set for $j_{100} \vee b_{-1}$ in w.

[10] Since 'Everybody is happy$_{19}$' does not obtain in w its intrinsic value in w is, according to clause (b) of IVO, equal to the least intrinsic value it has in any world in which it obtains. This value is 19, since this is its value in a world where there exists only one person, and that person is happy$_{19}$.

[11] The only minimal set for 'Nobody is happy$_{19}$' in w includes all basic states obtaining in w during the relevant time-interval. Hence, IV_w('Nobody is happy$_{19}$')=20.

[12] The last value-relation holds since there are worlds where a minimal set for 'Nobody is happy$_{20}$' includes, for example, a million basic states, each stating that someone is happy$_{21}$. Note that IV* cannot be repaired by including a clause to the effect that p is not intrinsically better than q if there are worlds where p is worse than q is in any world. This would lead to contradictions. For example, it would then be

In fact, Oldfield acknowledges also a third sense of intrinsic value, viz., 'absolute' intrinsic value.[13] He does very little to explicate this concept, but perhaps he would say that the absolute intrinsic value of p is equal to the (world-relative) intrinsic value of p in a world where p does not obtain, as given by clause (b) of IVO.[14] However, Oldfield claims that the notion of world-relative intrinsic value has an 'intuitive basis', and is therefore of interest independently of its relation to absolute intrinsic value. As an argument for this claim, Oldfield asks us to consider what 'our first, unsophisticated response' would be, if asked about the intrinsic value of a disjunction like $j_1 \vee b_{-20}$. He thinks that "one such response would be to say, 'it all depends on whether j_1 or b_{-20} is true'. In other words, there is an inclination to say that the value of $[j_1 \vee b_{-20}]$ does vary from world to world." (Oldfield (1977), p. 247.)

To me, this answer seems rather to reflect a failure to grasp the concept of intrinsic value. The *intrinsic* value of a state cannot depend on what other states happen to obtain. If this supposition is given up, the distinction between intrinsic and extrinsic value seems to collapse. Hence, it would be a great advantage if we can find a theory that avoids world-relative intrinsic values, but nevertheless gives plausible answers respecting the intrinsic values of non-basic states. I shall make three such proposals, the second of which closely resembles what I have taken to be Oldfield's notion of absolute intrinsic value.

2.

My first proposal is very simple. It consists in the idea that the intrinsic value of any state is a function of the intrinsic values of the basic states it entails. As long as happiness is the only evaluatively relevant dimension involved, it is perhaps also natural to think that the proper function is simply addition.[15] We then have

IV1 The intrinsic value of a state p is equal to the sum of the intrinsic values of the basic states entailed by p.

According to IV1 disjunctions like $j_2 \vee j_{-5}$ and $j_2 \vee j_5$ lack intrinsic value. Concerning the first disjunction, I believe that there is a fairly good argument for this contention. Consider two worlds, w and w', where j_2 and j_{-5}, respectively, are the only basic states. Intuitively, there is no intrinsically *bad* state in w, and no intrinsically *good* state in w'. But if we assume that $j_2 \vee j_{-5}$ has either negative or positive intrinsic value, and that the intrinsic value of a state does not vary between

---

the case both that $IV^*(j_1 \vee j_{-100}) > IV^*(j_0)$, and that $\neg(IV^*(j_1 \vee j_{-100}) > IV^*(j_0))$. Of course, one might respond to these counter-examples to IV* by introducing yet another sense of 'intrinsically better', satisfying the clause just mentioned. But this would surely be an unwarranted multiplying of senses.

[13] Oldfield (1977), p. 240.

[14] This is suggested by his remarks on pp. 240 and 245.

[15] This assumption is far from unproblematic, however. For example, it seems to lead to 'the Repugnant Conclusion'. (See Parfit (1984), p. 388.)

worlds, there is either an intrinsically bad state in *w*, or an intrinsically good state in *w*′. Hence, it might be said, ‘mixed’ disjunctions like $j_2 \vee j_{-5}$, with one good and one bad disjunct, can hardly have intrinsic value.

When it comes to ‘purely good’ or ‘purely bad’ disjunctions, on the other hand, the results of IV1 are more disputable. For example, $j_2 \vee j_5$ entails that *some* good state obtains, but does not entail that some bad state obtains. Hence, it might be said, it is intrinsically good. For analogous reasons, $j_{-2} \vee j_{-5}$ is intrinsically bad. The same can be said about states like ‘Someone is happy$_2$’ and ‘Ten people are happy$_{-5}$’, which also lack intrinsic value according to IV1. I shall now suggest a solution that incorporates these intuitions.

Let us keep Oldfield’s concept of an irrelevant state, as defined above, and let us generalize his notion of a minimal set, by removing the reference to a world. Thus, a set *S* is a *minimal p-set* iff (i) every member of *S* is a basic or irrelevant state, (ii) the conjunction of the members of *S* is contingent and entails *p*, and (iii) no proper subset of *S* is such that the conjunction of its members is contingent and logically entails *p*. Further, a set *S* is a *p-set* iff (i) *S* is a minimal *p*-set or a union of minimal *p*-sets, and (ii) the conjunction of the elements in *S* is contingent. Finally, *Min*(*p*) (*Max*(*p*)) is the sum of the intrinsic values of the members of a *p*-set *S*, such that the sum of the intrinsic values of the members of *S* is at least as small (great) as the sum of the intrinsic values of the members of any other *p*-set. We may now state our second proposal:

IV2 (a) A state *p* is intrinsically good iff Min(*p*)>0, intrinsically bad iff Max(*p*)<0, and intrinsically neutral iff it is neither intrinsically good nor intrinsically bad,
(b) if *p* is intrinsically good IV(*p*)=Min(*p*), if *p* is intrinsically bad IV(*p*)=Max(*p*), and if *p* is intrinsically neutral IV(*p*)=0.

According to IV2, only states which are guaranteed to realize something on balance good (‘to rate the universe a plus’) are intrinsically good, and only states which are guaranteed to realize something on balance bad (‘to rate the universe a minus’) are intrinsically bad.[16] Moreover, the intrinsic value of a state corresponds, so to speak, to the amount of good or bad it is guaranteed to realize. Thus, for example IV(j_5 & b_{-2})=3, IV($j_1 \vee j_5$)=1, IV($j_{-1} \vee j_{-5}$)= −1, IV(‘Somebody is happy$_2$’)=2, and IV($j_{-1} \vee j_5$)=0.[17] These results are plausible.

[16] The idea that an intrinsically good or bad state is one which rates the universe a plus or a minus, respectively, is found in Chisholm and Sosa (1966), p. 245.

[17] There is only one j_5 & b_{-2}-set, viz. $\{j_5, b_{-2}\}$. Since the sum of the intrinsic values of the members of this set is 3, Min(j_5 & b_{-2})=3. The ‘worst’ $j_1 \vee j_5$-set is $\{j_1\}$, hence Min($j_1 \vee j_5$)=1. The ‘best’ $j_{-1} \vee j_{-5}$-set is $\{j_{-1}\}$, hence Max($j_{-1} \vee j_{-5}$)= −1. The ‘worst’ $j_{-1} \vee j_5$-set is $\{j_{-1}\}$, while the ‘best’ $j_{-1} \vee j_5$-set is $\{j_5\}$. Hence Min($j_{-1} \vee j_5$)= −1, and Max($j_{-1} \vee j_5$)=5. Each ‘Somebody is happy$_2$’-set is a unit set containing a basic state, stating that a certain person is happy$_2$. Hence Min(‘Somebody is happy$_2$’)=2.

On the other hand, we would perhaps wish to maintain, for instance, that IV($j_0 \vee j_{100}$) > IV($j_0 \vee j_{-100}$), or that IV($j_{100} \vee j_{-1}$) > IV($j_1 \vee j_{-100}$). But IV2 entails that the intrinsic value of all these states is zero. Now, it might seem impossible to retain both the intuition that only those states that are guaranteed to realize some good or bad are intrinsically good or bad, and the opinion that IV($j_0 \vee j_{100}$) > IV($j_0 \vee j_{-100}$), or that IV($j_{100} \vee j_{-1}$) > IV($j_1 \vee j_{-100}$). However, both these ideas can be upheld, if we are prepared to accept that there are states which are intrinsically 'indeterminate', i.e., neither good nor bad nor neutral, and that some such states are nevertheless comparable with respect to intrinsic value. We can make room for indeterminate states in our theory by letting intrinsic values be represented by numerical intervals, rather than by single numbers. My final proposal, then, is:

IV3 (a) For any state p, IV(p) is equal to the interval Min(p) — Max(p),[18]
(b) p is intrinsically good iff Min(p)>0, intrinsically bad iff Max(p)<0, intrinsically neutral iff Min(p)=0 and Max(p)=0, and intrinsically indeterminate iff it is not intrinsically good, intrinsically bad, or intrinsically neutral,[19]
(c) if Min(p) ≥ Min(q) and Max(p) ≥ Max(q), then IV(p) ≥ IV(q),
(d) if IV(p) ≥ IV(q) and IV(q) ≥ IV(p), then IV(p) = IV(q),
(e) if IV(p) ≥ IV(q) and not IV(q) ≥ IV(p), then IV(p) > IV(q),
(f) if neither IV(p) ≥ IV(q) nor IV(q) ≥ IV(p), then IV(p) and IV(q) are incomparable.

IV3 entails that IV($j_0 \vee j_{100}$) > IV($j_0 \vee j_{-100}$), and that IV($j_{100} \vee j_{-1}$) > IV($j_1 \vee j_{-100}$).[20] Thus, it may correspond better than IV2 to certain intuitions. Moreover, since all four states are intrinsically indeterminate, these results do not violate the requirement that only states which are guaranteed to realise something on balance good or bad should come out as intrinsically good or bad.

Also concerning the evaluation of states like $j_2 \vee b_2$ and 'Someone is happy$_2$', IV3 differs from IV2. According to IV2, both these states have the intrinsic value 2. IV3, on the other hand, entails that IV($j_2 \vee b_2$) equals the interval 2 — 4,[21] and that

---

[18] If Min(p) = Max(p), I take the interval Min(p) — Max(p) to equal Min(p) and Max(p).

[19] Alternatively, one could somewhat weaken the criteria for intrinsic goodness and badness: (b') p is intrinsically good iff Min(p)≥0 and Max(p)>0, and intrinsically bad iff Max(p)≤0 and Min(p)<0. Adopting (b') means giving up the requirement that intrinsically good or bad states must guarantee some good or bad, respectively. On the other hand, (b') captures the intuition that p is intrinsically good (bad) if p is intrinsically better (worse) than any intrinsically neutral state.

[20] There are two 'best' $j_0 \vee j_{100}$-sets, viz. $\{j_{100}\}$ and $\{j_0, j_{100}\}$. Hence, Max($j_0 \vee j_{100}$)=100. Because $\{j_0\}$ is the only 'worst' $j_0 \vee j_{100}$-set, Min($j_0 \vee j_{100}$)=0. Analogously, Min($j_0 \vee j_{-100}$)= −100, Max($j_0 \vee j_{-100}$)=0, Min($j_{100} \vee j_{-1}$)= −1, Max($j_{100} \vee j_{-1}$)=100, Min($j_1 \vee j_{-100}$)= −100, and Max($j_1 \vee j_{-100}$)=1.

[21] The 'worst' $j_2 \vee b_2$-sets are $\{j_2\}$ and $\{b_2\}$, while the 'best' $j_2 \vee b_2$-set is $\{j_2, b_2\}$. Hence, Min($j_2 \vee b_2$)=2 and Max($j_2 \vee b_2$)=4.

IV('Someone is happy$_2$') equals the interval 2 — n×2, where n is the number of people in the most crowded possible world in which everybody is happy$_2$.[22] In these cases too, IV3 seems to yield as least as plausible results as IV2. There is, for example, a 'realisation' of $j_2 \vee b_2$, viz., j_2 & b_2, which is better than each of j_2 and b_2. Hence, one can reasonably claim that $j_2 \vee b_2$ is intrinsically better than j_2 and b_2.[23]

Yet another difference between IV2 and IV3 concerns negations, such as $\neg j_1$. While IV2 entails that IV($\neg j_1$)=0, IV3 yields that IV($\neg j_1$) equals the entire interval –n — n, supposing that there is an n such that –n and n are the greatest possible degrees of unhappiness and happiness, respectively.[24] Thus, $\neg j_1$ is an intrinsically indeterminate state, to an extreme degree. IV3 is therefore consistent with the claim I made above that negative states should not have a *determinate* (non-zero) intrinsic value. Nevertheless, it may appear somewhat counter-intuitive that the intrinsic value of $\neg j_1$ is incomparable to the values of most other (non-negative) states. We might be inclined to hold, for example, that IV(b_{10}) > IV($\neg j_1$). But when we consider the fact that $\neg j_1$ has both very good 'realisations', such as j_{100}, and very bad ones, such as j_{-100}, the conclusion that IV(b_{10}) and IV($\neg j_1$) are incomparable seems acceptable.

IV3 has a certain similarity to a proposal by Sven Danielsson, although he develops his theory in terms of sets of numbers, instead of intervals. His idea is roughly equivalent to the claim that p is intrinsically better than q iff Min(p) > Max(q). (Danielsson (1978), p. 11.) In my opinion, this suggestion is somewhat too 'guarded'. For example, it makes the states $j_1 \vee j_2$ and $b_1 \vee b_2$ mutually incomparable, with respect to intrinsic value, although they seem to be equally good. Likewise, the states $j_0 \vee j_{100}$ and $j_0 \vee j_{-100}$ are incomparable, according to Danielsson's theory. As I said above, it seems reasonable to hold that the former state is intrinsically better than the latter.

While Danielsson's proposal, like IV3, turns on the idea of assigning indeterminate intrinsic value to certain non-basic states, Michael Zimmerman has suggested that for example disjunctive states do not have any intrinsic value at all. He nevertheless agrees that a state such as $j_{10} \vee b_{10}$ is in a certain sense better than, say, j_{-10}. More precisely, $j_{10} \vee b_{10}$ is likely to have a 'derivative value simpliciter' which is higher than the intrinsic value of j_{-10}. The concept of derivative value

[22] If there are worlds with infinitely many simultaneously existing people, n is infinite. This might cause trouble for IV3, but problems with infinite numbers seem to confront all theories of intrinsic value. Even if no possible world contains infinitely many *simultaneous* people, there are presumably worlds with infinitely many people (or other sentient beings) spread out through (an infinite) time. Concerning states like 'Someone is happy to degree 2 during some time-interval' one might avoid these problems by denying that such states have intrinsic value. But this manoeuvre is hardly plausible when it comes to, for example, infinite conjunctions of basic states.

[23] The strength of this argument, which was pointed out to me by Krister Bykvist, is my reason for defining Max(p) and Min(p) in terms of 'best' and 'worst' p-sets, rather than directly in terms of 'best' and 'worst' *minimal* p-sets. Were Min(p) and Max(p) defined in the latter way, IV3 would, just like IV2, entail that IV($j_2 \vee b_2$)=2 and IV('Someone is happy$_2$') =2.

[24] The 'worst' $\neg j_1$-set is $\{j_{-n}\}$ and the 'best' $\neg j_1$-set is $\{j_n\}$. Hence, Min($\neg j_1$)= –n, while Max($\neg j_1$)=n.

simpliciter is defined like this: if (i) a state p has no intrinsic value, (ii) a state q has an intrinsic value of n, (iii) p obtains basically in virtue of q, and (iv) there is no other state r, such that p obtains basically in virtue of r, then p has a derivative value simpliciter of n. (Zimmerman (1983), p. 217.) A state p obtains in virtue of a state q iff (i) q is different from p, (ii) q entails p, and (iii) q obtains. A state p obtains *basically* in virtue of q iff (i) p obtains in virtue of q, and (ii) there is no state r, such that p obtains in virtue of r and r obtains in virtue of q. (Zimmerman (1983), p. 216.)

A problem with this idea, I think, is that it enables us to say that $j_{10} \vee b_{10}$ is better than j_{-10} only if it obtains. I would claim that $j_{10} \vee b_{10}$ is better than j_{-10} even if it does *not* obtain. Anyway, Zimmerman's proposal does not seem to work as it stands. If j_{10} and b_{10} both obtain, $j_{10} \vee b_{10}$ does not have any derivative value simpliciter, according to Zimmerman's definition, since it obtains basically in virtue of both j_{10} and b_{10}. Yet we would surely want to say that $j_{10} \vee b_{10}$ is somehow better than j_{-10} also if j_{10} and b_{10} both obtain.

I shall end this paper by pointing out two features which are shared by IV1, IV2, and IV3. First, these theories violate what we might call 'the intermediacy principle', stating that a disjunction is never intrinsically better, nor intrinsically worse, than both its disjuncts.[25] In the case of IV1 this violation is trivial, since IV1 entails that all disjunctions which entail no basic state have zero intrinsic value. To see that IV2 and IV3 also violate the intermediacy principle, consider the disjunction $p=(j_1 \vee (b_1 \mathbin{\&} d_{-100})) \vee (b_1 \vee (j_1 \mathbin{\&} d_{-100}))$. IV($p$)=1, according to IV2 and IV3. (Since p is logically equivalent to $j_1 \vee b_1$, the only minimal p-sets are j_1 and b_1.) But for each disjunct q of p, IV(q)=0 according to IV2, while IV(q) equals the interval −99 — 1 according to IV3. Hence, p is an example of a disjunction which is intrinsically better than both its disjuncts.[26] This conclusion need not embarrass us. Oldfield has given a convincing argument against the intermediacy principle. Consider all conjunctions of the form $j_1 \mathbin{\&} b_i$, where the state that Bill does not exist is included among the b_i. Some of these conjunctions are intrinsically good, some are bad, and some are neutral. Let G, B, and N be the disjunction of the good, bad, and neutral conjunctions, respectively. It seems that neither N nor the disjunction $G \vee B$ is intrinsically good. If a disjunction cannot be intrinsically better than each disjunct, we may infer that the disjunction $N \vee (G \vee B)$ is not intrinsically good, either. But $N \vee (G \vee B)$ is necessarily equivalent to j_1, which is intrinsically good, and must therefore itself be intrinsically good.[27]

---

[25] This principle is assumed, for example, in Chisholm (1975), p. 299 [* p. 233 of this volume].

[26] An example of this structure is discussed by Lennart Åqvist, who attributes it to Sven Danielsson and Bengt Hansson. Åqvist wants to maintain the intermediacy principle, while avoiding the conclusion that at least one of p's disjuncts is intrinsically not worse than $j_1 \vee b_1$. He therefore rejects the rule of substitution for logical equivalents. (Åqvist (1968), pp. 263-5.) Roderick Chisholm, on the other hand, keeps the intermediacy principle as well as the substitution rule, and accepts the result that neither of p's disjuncts is worse than $j_1 \vee b_1$. (Chisholm (1975), pp. 305-7 [* pp. 237-39 of this volume].)

[27] Oldfield (1977), p. 245.

The second feature worthy of notice is that certain 'holistic' value-phenomena obtain, according to both IV1, IV2, and IV3. None of these theories satisfies this principle:

(*) If p, q, and r are logically independent, and IV(p)=IV(q), then IV(p & r)=IV(q & r).

To find a counter-example to (*), for all three theories, let $p=j_1 \vee b_2$, let $q=b_1 \vee j_2$, and let $r=j_0$. According to IV1, IV(p)=0 and IV(q)=0, according to IV2, IV(p)=1 and IV(q)=1, and according to IV3, IV(p) and IV(q) both equal the interval 1 — 3. All three theories entail, however, that IV(p & r)=2 while IV(q & r)=1.

On the other hand, if the scope of (*) is restricted to basic states, it is valid according to both IV1, IV2, and IV3. This could be seen as a drawback of these principles, since there are probably few plausible axiologies which allow such a simple way of aggregating the values of basic states. However, there is no reason to suppose that there is a principle of aggregation that is consistent with all, or even most, plausible axiologies. Hence, this aggregation problem cannot, it seems, be fruitfully discussed without making much more detailed axiological assumptions than the simple hedonistic ones made in this paper.[28]

REFERENCES

Åqvist, Lennart, 'Chisholm-Sosa Logics of Intrinsic Betterness and Value', *Noûs* 2 (1968), 253-270.

Chisholm, Roderick M., 'The Intrinsic Value in Disjunctive States of Affairs', *Noûs* 9 (1975), 295-308 [* pp. 229-39 of this volume].

Chisholm, Roderick M., and Ernest Sosa, 'On the Logic of "Intrinsically Better"', *American Philosophical Quarterly* 3 (1966), 244-249.

Danielsson, Sven, 'Another Version of the Harman-Quinn-Oldfield Theory of Intrinsic Value', mimeographed, Uppsala 1978.

Oldfield, Edward, 'An Approach to a Theory of Intrinsic Value', *Philosophical Studies* 32 (1977), 233-249.

Parfit, Derek, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

Quinn, Warren S., 'Theories of Intrinsic Value', *American Philosophical Quarterly* 11 (1974), 123-132.

Sobel, J. Howard, 'Utilitarianisms: Simple and General', *Inquiry* 13 (1970), 394-449.

Zimmerman, Michael J., 'Evaluatively Incomplete States of Affairs', *Philosophical Studies* 43 (1983), 211-224.

[28] I wish to thank Krister Bykvist, Erik Olsson, and an anonymous referee for their helpful comments on an earlier draft.

CHAPTER 30

S. DANIELSSON

HARMAN'S EQUATION AND NON-BASIC INTRINSIC VALUE

1.

Gilbert Harman (1967) argued that the standard notion of intrinsic value is characterised by the equation

$$(1) \qquad I(A) = \sum_{A \Rightarrow B} v(B)$$

where $I(A)$ is the intrinsic value of A, $v(B)$ is the basic intrinsic value of B and $\Rightarrow$ is the relation of entailment. The intrinsic value of an alternative equals the sum of the basic intrinsic values of the alternatives entailed by it.

Thus according to Harman, intrinsic value is a) additive and b) in a sense, derived from basic intrinsic value. I have argued elsewhere that the latter of these two assumptions may shed some light on some standard problems with the former.[1] Here I shall assume that basic intrinsic value is defined, and that the question is how far Harman's equation can be used to calculate non-basic intrinsic value.

This question was raised by Warren Quinn (1974), who modified Harman's idea in order to make it possible to speak of the intrinsic value also of A's which entail no basically evaluated B, but entail, for instance, disjunctions of such B's. Edward Oldfield (1977) argued that Quinn's modification did not work and suggested another. Erik Carlson (1997) suggested a third one. Here I roughly agree with Oldfield against Quinn, and with Carlson against Quinn and Oldfield, but in certain respects I disagree with the three of them.[2]

---

[1] Danielsson (1997)

[2] The first version of this paper was written in 1978. Erik Carlson (1997) made me reconsider some of its arguments. I am grateful to him, as well as to Jan Odelstad and Krister Bykvist, who pointed out several difficulties.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 371-378.
© 2005 *Springer. Printed in the Netherlands.*

2.

Harman says that what may have value is that such and such is the case. He explicitly rejects discussing what a phrase such as ‘that the cat is on the mat’ may refer to.[3] Quinn and Oldfield assume that the bearers of intrinsic value are propositions. They identify propositions with sets of possible worlds. I shall essentially follow them in this respect, though, like Carlson, I would like to say that *A*, *B*, *C*, etc. are (possible) states of affairs rather than propositions. In any case, they are assumed to be identified by that-phrases. I shall also assume that there is a basic set *U* (the tautologous state of affairs), and that the variables ‘*A*’, ‘*B*’, ‘*C*’, etc. refer to subsets of *U*. Hence $\Rightarrow$, the relation of entailment, will be identified with $\subseteq$, set inclusion, and we shall be able to speak of the negation $\neg A$ of a state *A*, which is the complement in *U* to *A*, the disjunction $A \cup B$, which is the union of *A* and *B*, and the conjunction $A \cap B$, which is the intersection. That states of affairs are thus identified with sets of possible worlds is not important for the argument; it is simply a convenient way to model entailment and the notions of disjunction, conjunction, etc.

More important than these technical details is the question of how the distinction between intrinsic value and *basic* intrinsic value should be understood.

Harman argued that the standard account of intrinsic value forces its adherents to assume that there is basic intrinsic value. For the standard account, according to Harman, says that the intrinsic value of *A* is the sum of the intrinsic values of what is entailed by *A*. But that does not work if, for instance, both *A* and *A and B* have intrinsic value, for then the standard account entails that the intrinsic value of *A and B* is the intrinsic value of *A and B* plus the intrinsic value of *A*.

The force of this argument be what it may, the notion of basic intrinsic value obviously corresponds to the notion of *intrinsic value as a whole* in *Principia Ethica*,[4] where this is contrasted with *intrinsic value on the whole*, which is what is usually meant by ‘intrinsic value’. It is not all that easy, however, to provide a satisfactory definition. One might try with *underived* value, saying that the basic value of *A* is not dependent on the value of something else. This seems satisfactory in many cases. Thus, for instance, the hedonist would say that the value of the compound alternative *that there are happy egrets and there are stones* is not basic, since its value is derived from the value of the two conjunction members *that there are happy egrets* and *that there are stones.* And we could all agree that extrinsic value is not basic since it is, by definition, value as a means to an end, and, as such, dependent on the value of the end.

---

[3] We should perhaps distinguish between ‘It is good that the cat is on the mat’ and ‘It would be good if the cat were on the mat’, the former entailing, implying, presupposing or at least suggesting, that the cat is actually on the mat, the latter that it is not. The ascription of value to the cat’s perhaps just possibly being on the mat, however, seems to be the same in both cases, and this is what we are interested in here, although there seems to be no natural expression for it in ordinary language.

[4] Cf. G. E. Moore *Principia Ethica* (Cambridge: Cambridge University Press, 1903), section 129.

Some would like to say, however, that the negative value of *Alec is enjoying the thought that somebody is in great pain* is basic, although it is dependent on the negative value of the alternative *that somebody is in great pain.* Perhaps we could say that the characteristic feature of basic value is that it is not *entailed* by the value of something else. For it seems natural to say that while the hedonistic ascription of value to the conjunction above is entailed by the ascription of value to the conjunction members, the ascription of negative value to Alec's enjoying the suffering of somebody else is not entailed by the ascription of negative value to the suffering. In the latter case, the dependence is moral, not conceptual.

As the hedonistic example above illustrates, not every intrinsic evaluation has to be basic. It is less clear whether any basic evaluation can be non-intrinsic. To say that something is intrinsically good is to say that it has a positive value which is independent of the actual circumstances and consequences. To say that something is good in a non-intrinsic sense, then, would be to argue that it is good because of some presumed fact about the circumstances or consequences of the evaluated alternative. It is not obvious that you cannot base your value-argumentations ultimately on premises of that form.

But I am not concerned here with the question of whether every evaluative system must contain some basic intrinsic evaluation. I am only trying to clarify the distinction between basic and non-basic intrinsic value. The question is complicated by the fact that, as far as the Harman equation is concerned, the intrinsic value of a basic alternative need not be the same as the basic value of the alternative. In Quinn's, Oldfield's and Carlson's theories, this possibility is blocked; therefore, they can, and indeed do, drop the function v from Harman's approach. They need only speak of a set of *basic alternatives*, which are those basically assigned intrinsic value, and this is, in fact, the approach they take.

These authors can reduce the distinction between basic and non-basic intrinsic value to the difference between basic and non-basic alternatives because they assume that no basic alternative entails any other. If, for instance, *A* and *A and B* are both basic alternatives, and we have just one kind of intrinsic value, we are back to the situation which inspired Harman's original argument. Unfortunately, there are easily described evaluative systems with this particular feature. A retributivist with hedonistic inclinations, for instance, may argue that the whole composed by a criminal's committing a crime and his suffering punishment for the crime has a positive basic value, while the parts, the committing of the crime and the suffering of punishment, both have a negative basic value which is so great that the total intrinsic value of the whole is negative. And an egalitarian hedonist may argue that what has basic intrinsic value (positive or negative) is either that a certain individual is happy (or unhappy) to a certain degree, or that there is a certain difference between the degree of happiness of one individual and that of another. (In fact, it seems to me that no axiology with any plausibility at all can assume that no alternative with basic intrinsic value entails any other such alternative.)

Thus if we want a more general theory of intrinsic value we seem to need the notion of basic intrinsic value as a basic notion.

3.

Let us assume then that there is a set *U*, the subsets *A*, *B*, *C*,... of which are the states whose intrinsic value is at stake.

We assume that **V** is the set of basically evaluated alternatives, the alternatives for which the real-valued function v is defined.[5] To ensure that the sum of basic value mentioned in the Harman equation is well-defined, we assume that every non-empty (consistent) conjunction of members of **V** is finite.[6]

The Harman equation (1) now assigns positive or negative intrinsic value to an alternative if and only if it entails some member of **V**. On the one hand, this seems to be too restrictive. There may be non-basic alternatives which we would like to say are intrinsically better than certain others, and some alternatives which we would like to say are intrinsically good or intrinsically bad although they entail no member of the set **V**. It seems plausible to assume, for instance, that a disjunction $A \cup B$, where *A* and *B* both have a positive basic value, has itself a positive value, even if $A \cup B$ entails no member of **V**.

On the other hand, an alternative may also entail a member of **V**, and thus get a definite intrinsic value by the equation, although this value obviously differs from what seems to be intuitively plausible. Using hedonism as an illustration, assuming that pleasure and pain can, in some way or other, be measured, and that basic intrinsic value is assigned only to the precise degree of pleasure or pain of an individual, we may consider the alternative *that Alec is enjoying himself to degree 2 and Bertie is suffering to degree -1 or suffering to degree -10*. The Harman equation would give that alternative the intrinsic value of *Alec enjoying himself to degree 2*, since that is the basic value entailed. But clearly the compound alternative should be worse.

Hence the Harman equation in its original form must have a very restricted field of application. It may be applicable to members of **V**, conjunctions of such members, and alternatives formed by conjoining to such conjunctions alternatives which neither entail, nor are entailed by, members or conjunctions of members of **V**. In particular, the equation seems to be applicable to the total intrinsic value of a possible world, which hence would be the sum of the basic intrinsic values of alternatives that are actually realised in the world in question. This fits well with the ingenious dictum in Chisholm and Sosa (1964) that 'what is intrinsically good is what rates the universe a plus'. But Harman's main idea could be made more generally applicable, and this is what Quinn, Oldfield and Carlson have tried to do. Here is an alternative way of doing it.

---

[5] We usually consider the evaluations from an external perspective. We might as well take an internal one and say instead that **V** is the set of basically *valuable* alternatives.

[6] Notice that this does not mean that **V** is finite. There might well be an infinite number of levels of intrinsic value. But it means that there is a finite number of relevant aspects of the value of an alternative, in the sense that the intrinsic value of an alternative can always be accounted for by reference to a finite number of entailments.

4.

Compared to several earlier theories of intrinsic value, Quinn's theory seems to take a considerable step forward by assuming that an alternative may have an *indefinite* intrinsic value. Much logic of preference would have been happily undone if we had generally recognised this possibility. Such alternatives, however, are assumed by Quinn and Oldfield to have a definite *intrinsic value at a world* that shifts from one possible world to another. As Carlson emphasises, this seems to contradict a standard assumption about intrinsic value; the version of Harman's theory sketched below will *not* assume that intrinsic value varies from one possible world to another. The idea that intrinsic value may be indefinite, however, will be retained, and, like Carlson, we shall let the value of such an alternative be represented by a *set* of numbers instead of by a number.[7] The basic idea is simply that if A and B are both basic, then the value of $A \cup B$ is (represented by) the set $\{v(A), v(B)\}$.

Since we have assumed that basic alternatives may entail each other, this idea has to be somewhat restricted, however, lest it entail that the value of a basic alternative A is represented by a set containing every $v(B)$, where B is a basic alternative entailing A. In order to avoid this, we define an extended basic value-function v^*, which assigns a set of numbers to every member of the union closure of **V**, the set of all finite unions of members of **V**, which we call **V***:

(2) $v^*(A) = \{v(B)$*: B is in* **V**, $B \subseteq A$ *and there is no C in* **V** *such that* $B \subset C \subseteq A\}$.

Thus if the function v assigns a definite basic intrinsic value to the members of **V**, then v^* assigns an indefinite basic intrinsic value to disjunctions of these basic alternatives (including the one-term disjunctions of the basic alternatives themselves). The basic value brought into the world by *A or B* being the case is the basic value of A or the basic value of B. Notice that the restriction in (2) makes this basic idea applicable only to true disjunctions, so to speak, for if B is included in A, and $A \cup B$ is hence actually identical to A, then $v^*(A \cup B)$ will contain just $v(A)$, not $v(B)$. That which already has a definite basic value does not need an indefinite one. Or, to put the matter differently, the value of an alternative turns into a more indefinite one only if the alternative itself turns into a more indefinite one.

We then say that *B is a maximally v-definite consequence of A* if and only if $A \subseteq B$ and either B is in **V** or (*i*) B is in **V*** and (*ii*) there is no C in **V*** such that $A \subseteq C \subset B$.

It is evident that when we have listed the maximally v-definite consequences of A, then we have listed, as precisely as possible, which basically valuable things will

[7] Carlson actually uses *intervals* of real numbers. Since intervals can be taken to be sets of numbers, this may be considered a special case of the set-approach.

necessarily be in the world if *A* is true. And we take their added value to be the intrinsic value of *A*. Thus our new version of Harman's equation will be:

(3) $I(A) = \sum v^*(B)$

B is a maximally v-definite consequence of A.

We extend the relation > to sets of numbers, assuming that $v^*(A) > v^*(B)$ *if and only if $x>y$ for every x in* $v^*(A)$ *and y in* $v^*(B)$. We make the analogous assumption about the function *I*, and we make it possible to add indefinite values by assuming that $v^*(A) + v^*(B)$ is the set of all sums $x+y$ where x is in $v^*(A)$ and y in $v^*(B)$.

We assume further that *A* is intrinsically better than *B* if and only if $I(A) > I(B)$, and that *A* is intrinsically good (bad) if and only if $I(A)>0$ ($I(A)<0$). Given this, it seems natural also to assume:

(4) *If* $v(A)>0$*, then not:* $v(\neg A)>0$*; if* $v(A)<0$*, then not:* $v(\neg A)<0$*.*

(5) $v(U) = 0.$

If it is good (bad) that *A*, then it is not also good (bad) that *not A*. To be good (bad) is to be better (worse) than the tautology.[8]

It is evident that this version satisfies the demands on the treatment of disjunctions which were formulated at the end of section 3. Let us assume that the basic intrinsic value of *Alec is enjoying himself to degree 2* is 2, the basic intrinsic value of *Bertie is suffering to degree -1* is -1, and the basic intrinsic value of *Bertie is suffering to degree -10* is -10. The maximally v-definite consequences of *Alec is enjoying himself to degree* 2 *and Bertie is either suffering to degree -1 or suffering to degree -10* are (i) the tautology, (ii) *Alec is enjoying himself to degree 2*, and (iii) *Bertie is suffering to degree -1 or Bertie is suffering to degree -10*; the v*-values of these are {0}, {2} and {-1, -10}, and hence the intrinsic value of the alternative at stake is {1, -8}.

5.

This version is simpler than those of Quinn, Oldfield and Carlson. It is also more general, since it does not assume that the basic alternatives are logically independent of each other. Apart from these general features, the most striking particular deviations from Carlson's theory are probably that in the present one *every* member of $I(A)$ has to be greater than *every* member of $I(B)$ for *A* to be better than *B*, and that the value of *A and B* has no particular role in the calculation of the value of *A or B.*

[8] It certainly may be questioned whether the value of the tautology is really comparable to the value of anything else. But the main point of these assumptions is to make the numerical machinery work more efficiently.

Regarding the former, Carlson presents a couple of examples intended to show that the version chosen here is not a reasonable one. In the most important case, we are invited to compare *A or B* to *B* when *A* is better than *B* (and *A* and *B* are both basic). Is it not obvious that, since *A or B* involves the possibility of *A*, which is better than *B*, that the disjunctive alternative is also a better one? I do not think so. Perhaps it could be argued that, given some kind of choice situation in which we in some sense have to choose between *A or B* and *B* (it is not very clear how such a situation should be described), then choosing *A or B* is a preferable choice. But it is not obvious that 'intrinsically better than' should be equated with the formulation in terms of choice.

And if we assume that *A or B* is better than *B* in case *A* is better than *B* (and *A* and *B* both are basic), then our theory of intrinsic value is hardly compatible with hedonism.[9] For hedonism, strictly speaking (or perhaps we should say one strict kind of hedonism), is the assumption that *A* is intrinsically better than *B* if and *only* if *A* contains, that is, entails, a greater balance of pleasure over pain than *B* does. And it is obvious that *Alec is enjoying himself or Bertie is suffering* does *not* entail a greater balance of pleasure over pain than does *Bertie is suffering.* Perhaps we could say that the former alternative *might* bring positive value into the world by being true (*might* render the Universe a plus), while the latter certainly will not. But hedonism as it is usually conceived assigns positive value to pleasure, not to possible pleasure. As far as *Alec is enjoying himself* and *Bertie is suffering* are concerned, there is more pleasure in the world if the former is true than if the latter is true. But it is not the case that, as far as *Alec is enjoying himself or Bertie is suffering* and *Bertie is suffering* are concerned, there is more pleasure in the world if the former is true than if the latter is.[10]

It also seems to me that if, when *A* is better than *B*, we do not say that *A or B* is *better than B*, but say instead simply that *A or B* is *at least as good as B*, then we get a terminology which makes it possible to make some finer, though still very simple, discriminations. I think we should consider the possibility that *A* is intrinsically at least as good as *B* if and only if $x \geq y$ for every x in $I(A)$ and y in $I(B)$. That means, of course, given the assumption of 'better than' above, that we cannot adopt the equivalence between *A is better than B* and *A is at least as good as B and B is not at least as good as A*, which so often appears in the literature. But the equivalence seems to be a necessary one only if the betterness-ordering is a weak order. Nor will *A is at least as good as B* be equivalent to *A is better than B or equal in value to B.*

The notion of equality in intrinsic value actually appears to be a much more problematic one than the notion of betterness. There seem to be a number of more or

---

[9] Some people may be inclined to say that if our theory of intrinsic value is incompatible with hedonism, then so much the worse for hedonism. But although I am sure that hedonism is wrong, I do not think that it is wrong for logical reasons.

[10] We certainly can imagine a sort of hedonism which assigns basic intrinsic value also to the disjunction. Then the disjunction might be intrinsically better than the second disjunction member. So this kind of hedonism is compatible with our theory of intrinsic value, while what here is taken to be hedonism proper is not compatible with the alternative theories.

less plausible candidates ranging from *neither better nor worse than* and upwards in logical strength. One of the best, perhaps, says that *A* has the same intrinsic value as *B* when something is intrinsically better than *A* if and only if it is intrinsically better than *B*, and intrinsically worse than *A* if and only if it is intrinsically worse than *B*.[11] And to incorporate this idea into our theory smoothly, we can say that *A is equal in intrinsic value to B if and only if the maximal member of I*(*A*) *equals the maximal member of I*(*B*) *and the minimal member of I*(*A*) *equals the minimal member of I*(*B*). (The present theory will hence, on this point, agree with Carlson's.)

As to the role of the value of *A and B* in the evaluation of *A or B*, Carlson argues, like Oldfield, that in some cases *A or B* is better than *C* because *A and B* is better than *C*. For some reason, they assume that among all subalternatives of *A or B* precisely the subalternative *A and B* (if that is consistent) should have a special standing. Carlson argues that *A and B* is a 'realisation' of *A or B*, implying that the value of an alternative is dependent on the value of its 'realisations'. But it is in no way clear why *A and B* is a 'realisation' of *A or B* in a sense different than for any subset of *A or B* whatsoever.

REFERENCES

Carlson, Erik: 'The Intrinsic Value of Non-Basic States of Affairs'. *Philosophical Studies* 85 (1997): 95-107 [* pp. 361-70 of this volume].

Danielsson, Sven: 'Harman's Equation and the Additivity of Intrinsic Value'. In Lindahl, Needham, Sliwinski: *For Good Measure: Philosophical Essays Dedicated to Jan Odelstad on the Occasion of his Fiftieth Birthday* (*Uppsala Philosophical Studies*, 46 (1997)): 23-34.

Halldén, Sören: *On the Logic of Better*, Library of Theoria 1957.

Harman, Gilbert: 'Toward a Theory of Intrinsic Value'. *Journal of Philosophy* 64 (1967): 792-804 [* pp. 349-60 of this volume].

Oldfield, Edward: 'An Approach to a Theory of Intrinsic Value'. *Philosophical Studies* 32 (1977): 233-49.

Quinn, Warren: 'Theories of Intrinsic Value'. *American Philosophical Quarterly* 11 (1974): 123-32.

[11] This idea, though not for *intrinsic* betterness, is used already in Halldén (1957).

CHAPTER 31

F. FELDMAN

BASIC INTRINSIC VALUE

1. SOME PUZZLES IN AXIOLOGY

When called upon to tell us what hedonism is, moral philosophers often start out by contradicting themselves. An extreme and blatant case would look like this:

> Hedonism: the view that (i) pleasure is the only thing that is intrinsically good, and (ii) pain is the only thing that is intrinsically bad; furthermore, the view that (iii) a complex thing such as a life, a possible world, or a total consequence of an action is intrinsically good iff it contains more pleasure than pain.

The contradiction seems obvious, but since so many philosophers evidently don't see it, perhaps it will be useful to employ the sledgehammer. According to the first clause, pleasure is the only thing that is intrinsically good. This implies that nothing distinct from pleasure is intrinsically good. Since pleasure (whatever precisely it may be) cannot be identified with any life or with any possible world, it would follow that no life or possible world is intrinsically good. Yet, according to the third clause, complex things such as lives and possible worlds are also intrinsically good when they contain a favorable balance of pleasure over pain.

You cannot have it both ways: either (a) pleasure is the only thing that is intrinsically good (in which case no lives and no possible worlds are intrinsically good), or else (b) some lives and some worlds are intrinsically good (in which case pleasure is not the only thing that is intrinsically good).

I claimed that many philosophers have contradicted themselves in this way. I will cite two prominent examples.

In *Brentano and Intrinsic Value*, Roderick Chisholm seems to commit himself to this contradiction. In an attempt to sketch a simple form of hedonism, Chisholm says, 'States of pleasure are the only things that are intrinsically good and states of

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 379-400.
© 2005 *Springer. Printed in the Netherlands.*

displeasure or pain are the only things that are intrinsically bad.'[1] He goes on to specify the intrinsic values (according to the same theory of value) of various "mere sums" and other complex states of affairs including organic unities of various sorts. Some of these are compound states of affairs consisting of one person feeling some pleasure and another feeling some pain. Clearly, none of the complexes under consideration is a "state of pleasure" or a "state of pain" as Chisholm understands these concepts. Yet Chisholm says that some of these compounds have positive intrinsic value and so are intrinsically good. Thus, Chisholm states hedonism in a self-contradictory way, saying at the outset that pleasures are the only intrinsic goods and saying shortly thereafter that other things are also intrinsically good.

Richard Brandt seems equally guilty. In his article, 'Hedonism' in *The Encyclopedia of Philosophy*, he describes hedonism as the view that:

> ...only pleasant states of mind are desirable in themselves; that only unpleasant states of mind are undesirable in themselves; and that one state of affairs is more desirable in itself than another state of affairs if and only if it contains more (in some sense) pleasant states of mind than the other...[2]

Surely it is wrong to say both that pleasant states of mind are the *only* things that are desirable in themselves, and also that *other* states are desirable in themselves if they contain more pleasant states.

The first puzzle in axiology concerns this contradiction. How are we to avoid it? I will claim that an appropriate appeal to "basic intrinsic value states" would help to avoid this sort of contradiction, and thus to solve the first puzzle. Somewhat more substantially, I will suggest that the appeal to basic intrinsic value states will facilitate the formulation of any axiological theory.

The second puzzle concerns one sort of extrinsic value – value as a means. Since the time of Plato and Aristotle, moral philosophers have recognized a distinction between intrinsic value and this sort of extrinsic value.[3] Let us agree provisionally that something is intrinsically good if it is good "as an end", or "in itself", or "for its own sake", or "in virtue of its own nature". How then shall we explain value as a means (or "instrumental") value?

The standard answer is that something is good as a means if it causes something that is intrinsically good. Moore suggests something like this in *Principia Ethica*, where he says:

> Whenever we judge that a thing is 'good as a means', we are making a judgment with regard to its causal relations; we judge *both* that it will have a particular kind of effect, *and* that that effect will be good in itself.[4]

C. I. Lewis says a similar thing in *An Analysis of Knowledge and Valuation*:

[1] Roderick Chisholm, *Brentano and Intrinsic Value*, 74 [* p. 308 of this volume].
[2] Richard Brandt, 'Hedonism', 432.
[3] See, for example, *Republic* 357; *Nicomachean Ethics*, 1096b13.
[4] G.E. Moore, *Principia Ethica*, 22.

> A thing *A* will never be said to have extrinsic value or instrumental value, unless it is meant to imply that there is some other thing, *B*, to which it is or may be instrumental, which has intrinsic value.[5]

In his description of what he takes to be "the standard account", Gilbert Harman introduces this formula:

> S is instrumentally desirable to the extent that S does not entail but "leads to" something intrinsically desirable.[6]

The general idea here seems to be that a thing has instrumental value if it causes something with intrinsic value. This is typically conjoined with the further assumption (clearly suggested by Harman's formula) that the instrumental value of a thing is equal to the intrinsic value of something that it causes. Since instrumental values vary from world to world, depending upon what a thing happens to cause in a world, anyone adopting this approach should be careful to relativize instrumental values to worlds. Making use of the idea that the bearers of value are all states of affairs, we can put this point by saying:

> EV1: The instrumental value of a state of affairs, p, at a world, w, is some number, n, iff p causes some state of affairs, q, at w, such that the intrinsic value of q is n.

This has strange implications. Suppose p causes an intrinsically good thing with value +10 and also causes an intrinsically bad thing with value –10 at w. Then EV1 implies that the instrumental value of p is both +10 and –10; that p is both instrumentally good and instrumentally bad. That seems wrong. The natural revision is this:

> EV2: The instrumental value of a state of affairs, p, at a world, w, is the sum of the intrinsic values of all the things, q, such that p causes q at w.

EV2 gives each state of affairs a unique instrumental value at each world. Thus, nothing is both instrumentally good and instrumentally bad at the same world according to EV2. But EV2 has other problems. Consider a case in which Bob reads a delightful book. We may focus on this state of affairs:

> B: Bob reads at 9:00PM on Monday evening.

Suppose B makes Bob happy to degree +10 while he reads, and sits in his easy chair, and wears his eyeglasses. Suppose B never makes Bob or anyone else happy or

---

[5] C. I. Lewis, *An Analysis of Knowledge and Valuation*, 385.
[6] Gilbert Harman, 'Toward a Theory of Intrinsic Value', 796 [* p. 353 of this volume].

unhappy at any other time. In this case, B has many consequences, including the following:

C1: Bob being happy to degree +10 at 9:00PM on Monday evening.
C2: Bob being happy while reading.
C3: Someone being happy to degree +10.
C4: The man in Bob's easy chair being happy to degree +10.
C5: Bob being happy at least to degree +9.
C6: Bob being happy to degree +10 while sitting in his easy chair and wearing his eyeglasses at 9:00PM on Monday evening.
C7: Either Bob being happy to degree +10 or Babe being happy to degree +15.

If a certain simple and familiar form of eudaimonism is true, then each of these is "good in itself", or "good in virtue of its own nature". Some of these appear to have an intrinsic value of +10; others appear to have positive intrinsic value, though perhaps not +10. If so, and if B causes no bad things, then EV2 implies that the instrumental value of B is tremendously high. This is clearly wrong.

Notice that among the consequences of B there are a lot of states of affairs that somehow or other relate to the fact that Bob enjoyed 10 units of happiness at 9:00PM on Monday night. Some entail it, some are entailed by it, and some seem, in a vague way, to coincide with it. Cl seems to contain precisely the information that we are interested in – provided that we endorse the assumed form of eudaimonism – that is, neither more (like C6) nor less (like C2) nor muddled nor misleadingly incomplete. Suppose we had some way to disregard all of these other states of affairs and to count just Cl. Then we would have a way of explaining the instrumental value of B: it causes Cl, and the intrinsic value of Cl is +10. Therefore, we might say, the instrumental value of B is +10. However, what justification is there for taking account of the intrinsic value of Cl while ignoring the intrinsic values of all the other consequences of B? After all, each of them seems to be good in itself and each of them is just as much a consequence of B.

Thus, we have our second puzzle in axiology: how are we to explain the instrumental value of a state of affairs? I will claim again that an appropriate appeal to basic intrinsic value states will figure in the most plausible solution to the puzzle. Roughly, my suggestion will be that we understand basics in such a way that C1 is a basic intrinsic value state, while all the other consequences of B are not basics. Then we can say that the sum of the intrinsic values of the basics among the consequences of B is +10, and that for this reason B has an instrumental value of +10, which seems just right.[7]

[7] In Section 4 below, I discuss some further difficulties with this account of instrumental value and I sketch a way of dealing with some of them.

The third puzzle concerns the intrinsic values of complex things, such as lives, total consequences, and possible worlds. For purposes of illustration, let us focus on lives. Since the time of Plato and Aristotle, it has been assumed that one of the main aims of axiology is to tell us something about "the good life". An axiology is called upon to provide principles specifying the intrinsic value of each life. How is this to be done?

The life of a person at a world might be taken to be the conjunction of all the facts about the person at the world. We might say that a state of affairs is "about" a person if any sentence of the most appropriate sort for expressing that state of affairs has, as its grammatical subject, a bare name of the person. Then, we can say that the intrinsic value of the life of a person at a world is equal to the sum of the intrinsic values of the states of affairs about him that are true at that world.

We can formulate this as follows.

> LIV1: The intrinsic value of the life of S at world w = the sum of the intrinsic values of all states of affairs, q, such that q is about S and true at w.

I relativize to worlds since (as I see it) a given person might exist at several different possible worlds. His life at one world might be worth more or less than his life at another.

There are several apparent problems with LIV1. (a) Suppose S lives a tremendously boring life at w; he experiences just one instant of happiness, as indicated by the statement 'S is happy to degree +3 at time t'. Suppose no other resident of w has any happiness or unhappiness. Then LIV1 yields the result that S's life is infinitely valuable. For all of these are true and about S at w:

> S1: S is happy to degree +3 at t.
> S2: S is happy to degree +3 at t while 2 + 2 = 4.
> S3: S is happy to degree +3 at t while 3 + 3 = 6.
> S4: S is happy to degree +3 as a result of what happened to the chickens; etc.

(b) Suppose some other resident, Z, of this same world lives an even more boring life than S. He never experiences any happiness or unhappiness. Still LIV1 yields the result that Z's life is infinitely valuable. For each of these is true and about Z:

> Z1: Z is such that S is happy to degree +3 at t.
> Z2: Z is such that S is happy to degree +3 at t while 2 + 2 = 4.
> Z3: Z is such that S is happy to degree +3 at t while 3 + 3 = 6.
> Z4: Z is such that S is happy to degree +3 at t as a result of what happened to the chickens; etc.

Again, the natural solution seems to be that we should appeal to basic intrinsic value states and we should say that the value of S's life is equal to the sum of the intrinsic values of the basics that are true at w and appropriately about S. If we select our basics correctly, it will turn out that there is just one basic intrinsic value state that (a) is properly about S in the example cited above, and (b) is true, and (c) has an intrinsic value other than zero, and that is S1. Then we can let the intrinsic value of S's life be equal to the sum of the intrinsic values of these basics. In other words,

> LIV2: The intrinsic value of the life of S at world w = the sum of the intrinsic values of all states of affairs, q, such that q is a basic intrinsic value state; q is about S; and q is true at w.

Given my assumptions about basics, LIV2 correctly implies that the value of S's life is +3. In the case of Z, we will find that every basic intrinsic value state that is about him and true in this example has an intrinsic value of zero. It, therefore, turns out that LIV2 declares his life to be worthless, as our eudaimonistic axiology intends it to be.

Another traditional aim of axiology is the provision of principles about the evaluation of worlds. This is thought to be important for several reasons. For example, it plays a central role in many discussions of the problem of evil. It may also play a role in the formulation of certain views in the normative ethics of behavior.[8] Thus, there is some interest in finding a principle telling us the intrinsic value of a world. Again, we don't want to say:

> WIV1: The intrinsic value of a possible world, w, = the sum of the intrinsic values of all states of affairs, q, such that q is true at w;

but we might want to say:

> WIV2: The intrinsic value of a possible world, w, = the sum of the intrinsic values of all states of affairs, q, such that q is a basic intrinsic value state and q is true at w.

Again, if we have selected our basics in the way I am imagining, this will give positive intrinsic values to things like S1 and negative intrinsic values to things like

> J1: Jones is happy to degree –10 at t.

[8] In my *Doing the Best We Can* I present a normative theory according to which what we ought to do depends upon what happens in the "best accessible worlds". Clearly, this view makes essential use of the idea that worlds can be evaluated in terms of intrinsic value

If we select our basics correctly, and we assign intrinsic values to them correctly, it will be reasonable to say that the intrinsic value of a world is the sum of the intrinsic values of all the basics true there.

Similar considerations apply in the case of the total consequences of actions.

This gives us another reason to assume that there are basic intrinsic value states. It helps us to formulate principles that determine the intrinsic values of complex things such as consequences, lives, and worlds.

There are other reasons to introduce the concept of basic intrinsic value. Some of these have to do with the cluster of issues surrounding defeat and enhancement of intrinsic value and organic unities. They will be discussed later.

2. METAPHYSICAL DIGRESSION

A critic might raise the following criticism: 'You have assumed that the bearers of intrinsic value are states of affairs. States of affairs are very "fine-grained" entities. (Consider Cl-C6 above.) Once you make this assumption, you will need some way to avoid double counting, and maybe the appeal to basics will be unavoidable. However, if you had chosen "coarser" entities, such as concrete events, as the bearers of intrinsic value, the whole issue would have been avoided. There would have been no need for basics. So the puzzle is simply an artifact of your own metaphysics.'

I think it's important to see that this criticism is misguided. We will need to distinguish between basics and non-basics even if we assume that the bearers of value are concrete events.

Suppose that the bearers of value are concrete events. Suppose that these expressions all serve to pick out the same concrete event:

C1: Bob being happy to degree +10 at 9:00PM on Monday evening.
C2: Bob being happy while reading.
C3: Someone being happy to degree +10.
C4: The man in Bob's easy chair being happy to degree +10.
C5: Bob being happy at least to degree +9.
C6: Bob being happy to degree +10 while sitting in his easy chair and wearing his eyeglasses at 9:00PM on Monday evening.

We can say that there is just one concrete event involved here and that it has an intrinsic value of +10. The occurrence of this event at a world increases the value of the world by just 10 points. There may seem, therefore, to be no need (or even *possibility*) to distinguish the basics from the non-basics.

However, the appeal to concrete events does not solve the problem. Consider the concrete event, e1, that consists in Bob's being happy at 9:00PM on Monday evening. Perhaps each of Cl-C6 somehow indicates this single event. Now consider the event, e2, that is Bob's whole life. In light of the fact that e2 begins long before

e1, and ends long after e1, and contains many parts that e1 does not contain, we surely cannot identify e2 with e1. Yet if e1 and e2 are distinct and each has a positive intrinsic value, then it may seem that each contributes to the intrinsic value of the world. But to count them both is to engage in double counting, since the intrinsic value of e1 is already included in the intrinsic value of e2.

There are deeper problems about the move to concrete events. One of these turns on the fact that we may want to make a number of claims about the modal features of intrinsic value. For example, we may want to say that each thing has its intrinsic value of necessity. But there is no consensus about the modal features of concrete events. Consider the concrete event that consists in Bob's being happy to degree +10 at 9:00PM. Could it have happened a bit earlier in the evening? Could it have involved a slightly smaller amount of happiness? Could it have happened to Babe instead of to Bob? If the item in question is truly "concrete", I see no way to assure negative answers to these questions. Thus, it appears that a concrete event that has an intrinsic value of +10 might have had an intrinsic value of +9; a concrete event that serves to enhance the intrinsic value of Bob's life might have served to enhance the intrinsic value of Babe's life. It would be better to start with metaphysical assumptions that rule out such possibilities.

Thus, the move to concrete events (a) will not obviate the need for basics and (b) will introduce puzzles and problems of its own. So I prefer to stick with the assumption that the fundamental bearers of intrinsic value are very finely individuated states of affairs. Given that assumption, it seems reasonable to try to distinguish the basics from the non-basics.

3. GENERAL CHARACTERIZATION OF BASICS

Let us agree, then, that there might be some justification for the assumption that there are some selected states of affairs that are basic intrinsic value states ("basics"). One natural question at this point is this: which states of affairs are the basics?

But of course this question cannot be answered without begging many of the central substantive questions of axiology. Different axiological views entail different views about which states of affairs are the basics. What is a basic on one axiology might have no value, basic or otherwise, in another axiology.

Since I believe that the recognition of basics is useful no matter what our axiology, I want to give a general characterization of basics, one that will hold no matter what axiological theory we ultimately decide to accept.

I think there are six main features of basics.

1. *Every basic is a pure attribution of a core intrinsically valuable property or relation.* Ideally, an axiological theory should precisely identify the states of affairs that it takes to be the ultimate sources of intrinsic value. These are the basics on that

axiology. Thus, on the simple form of eudaimonism I mentioned earlier, the basics would be states of affairs such as the one expressed by this:

J12: Jones being happy to degree +12 at noon, March 25, 2000.

One crucial feature of J12 is that it is a "pure attribution" of *happiness.* The state of affairs attributes it directly to a person and not via any of his properties. The phrase used to express this state of affairs picks out Jones by the use of a meaningless "tag" and not by any meaningful description. The state of affairs attributes a precise degree of happiness (+12) to that person. It attributes it to him at a specific, named time.[9] It does nothing else.

I assume that some states of affairs are "directly about" certain people, times, and numbers. Thus, I assume that there is a state of affairs that says, with respect to Jones, now, and +12, that he is happy at it (the time) and to it (the number). Some philosophers of language might think that this state of affairs is composed of Jones, now, +12, and happiness. If we take this view (and I am not sure that I want to) we can say that happiness is the only "conceptual component" of the state of affairs. The other components of the state of affairs are "objects" such as Jones, now, and +12. For this reason, I will say that it is a pure attribution of happiness.

We can say in general that a state of affairs is a pure attribution of F iff there is something, x, such that the state of affairs is the state of affairs of x's having F (where x appears directly in the state of affairs). That covers the one-place case. For relations: p is a pure attribution of *x,y|Rxy* if there are individuals, x and y, such that p is the state of affairs of x's bearing R to y (again where x and y appear directly in the state of affairs); and so on for relations involving larger numbers of terms.

Thus, the basics of the simple form of eudaimonism are all pure attributions of *x,t,n|x is happy at t to n.* Since each such basic contains no other information beyond the mere identity of the recipient of the happiness, the time at which the recipient receives it, and the amount received, I will say that it contains no "extraneous information." It contains just the core information that a simple eudaimonist will think is relevant to intrinsic value.

Perhaps this helps to explain why the ancients persisted in saying that eudaimonism is the view that *happiness* is the sole intrinsic good. Such talk certainly appears to attribute intrinsic value to a property. But what they surely meant was that the intrinsically good basics are all pure attributions of happiness. Those who said that hedonism is the view that *pleasure* is the sole intrinsic good must similarly have meant that, according to hedonism, every intrinsically good basic was a pure attribution of pleasure.

---

[9] Some problems remain. If we assume that the temporal indicators refer to *instants* of time, and that there are infinitely many of these, each of which is unextended in time, we will run into trouble. Since even a short period of happiness will presumably involve happiness at infinitely many of these instants, it will be hard to find the sum of the values of these things. Perhaps a better approach involves the notion that time can be exhaustively broken down into a finite number of non-overlapping brief segments. The temporal indicators may then be taken to refer to these tiny intervals.

It also gives us a way of explaining what's distinctive about "pluralist" axiologies. It is not sufficient merely to say that pluralism is the view that there are several different sorts of intrinsically good states of affairs. This would make even the most hardcore monistic hedonism into a form of pluralism, since even in such an axiology the good states of affairs differ in a number of respects, e.g. some are about you, some are about me, some are about large amounts of happiness, and some are about small amounts of happiness. But the appeal to basics may help. We can say that an axiology is pluralist just in case it implies that some good basics are pure attributions of one property, and other good basics are pure attributions of another property. Thus, for example, if someone says that some good basics are pure attributions of pleasure and others are pure attributions of knowledge, then that person is a pluralist in axiology.

2. *Every basic has a determinate intrinsic value.* Eudaimonism of the sort imagined presupposes that every basic intrinsic value state has a determinate amount of intrinsic value. Our eudaimonist would say that the basics are all pure attributions of happiness. He could go on to say that the "H-value" of a basic is equal to the precise amount of happiness (positive or negative) attributed by that basic. Then he could say that the intrinsic value of each basic is equal to its H-value. Thus, given this eudaimonistic axiology and the imagined view about which states of affairs are the basics, it follows that every basic has a fully determinate, absolutely precise intrinsic value.

What's true of these basics on this eudaimonistic axiology is the universal feature of basic intrinsic value states: each of them has a perfectly determinate intrinsic value. But again, different axiologies identify different sets of basics. As a result, we have to be careful to state this thesis in a slightly more guarded way: *each axiology selects a set of basic intrinsic value states; for each such axiology, the basics it selects must have perfectly determinate intrinsic values.*

States of affairs that are not basic intrinsic value states may fail to have fully determinate intrinsic values. A number of philosophers have tried to formulate principles that determine the intrinsic value of each state of affairs (including the non-basics). They have struggled to identify the intrinsic values (given specified axiologies) of various states of affairs.[10] Philosophers in this tradition evidently assumed that, since eudaimonism implies that *Jones being happy about the circus* is good in itself, it must be good in itself to some particular degree.

But how intrinsically good is it?

In his 1967 paper, "Toward a Theory of Intrinsic Value",[11] Gilbert Harman introduced a general formula that seems to be relevant here. He proceeded by first

---

[10] See, for example, Chisholm's 'The Intrinsic Value in Disjunctive States of Affairs' [* pp. 229-39 of this volume] and Quinn's 'Theories of Intrinsic Value'.

[11] Throughout this paper I often have in mind things that Gilbert Harman said in his 1967 paper 'Toward a Theory of Intrinsic Value' [* pp. 349-60 of this volume]. I have to admit that I don't fully understand

considering a formula that does not work. It is this: 'S is intrinsically desirable to the extent that S entails something that is intrinsically desirable'[12] Harman's remarks are not entirely clear to me, but in light of what he later says, he might have been thinking of this:

IV1: IV(p) = the sum, for all q, such that p entails q, of IV(q).

Of course this is not acceptable, since if S is intrinsically good, it entails infinitely many other intrinsic goods. Then its intrinsic value goes through the roof. Harman makes some remarks that suggest that he would reject IV1 for just this reason.[13]

Then Harman says:

IV2: IV(p) = the sum, for all q, such that p entails q, of BIV(Q).

We may assume that BIV is a function that takes us to the intrinsic value of p if p is a basic intrinsic value state; otherwise, it takes us to zero. In this way, IV2 can be understood as the principle that the intrinsic value of any state of affairs is equal to the sum of the intrinsic values of the basics it entails; or, equivalently, equal to the sum of the basic intrinsic values of all the states of affairs it entails. This generates acceptable results in some cases. For example, (still assuming the form of simple eudaimonism that we have been using) consider:

N: S being happy to degree +12 at t1 & Z being happy to degree +10 at t2.

N entails exactly two basics and the sum of their intrinsic values is +22. IV2 implies that the intrinsic value of N is +22. That seems correct.

However the proposal is wrong in just about every other sort of case. Consider *someone is happy to degree +12*. It entails no basics, and so gets an intrinsic value of zero according to IV2. Yet according to the traditional way of thinking of intrinsic value, it is viewed as being intrinsically good. Consider *the man in the blue hat is happy to degree +12*. It also gets an intrinsic value of zero on IV2 and it too would traditionally be taken as good. Consider:

M: S being happy to degree +12 at t1 or Z being happy to degree +10 at t2.

M seems pretty good – it guarantees at least 10 units of intrinsic value – yet IV2 implies that it has an intrinsic value of zero. A wide variety of other examples fail to receive the desired intrinsic values on IV2.

---

every detail of his argument. Nevertheless, he comes to the conclusion that we should recognize basics, and he seems to make use of some considerations that are at least similar to things I will say.

[12] Harman, 'Toward a Theory of Intrinsic Value', 798 [* p. 353 of this volume].

[13] Harman, op. cit., 799 [* pp. 354-55 of this volume].

Another proposal makes the intrinsic value of S equal the sum of the intrinsic values of basics that "make it true" at w. Consider the disjunction M. If the first disjunct makes it true at w, then the intrinsic value of the disjunction is +12. If the second disjunct makes it true, then its intrinsic value is +10. If both disjuncts make it true, then its intrinsic value is +22. In the case of the man in the blue hat, if it's true, then some basic must make it true. Maybe it is *Smith being happy to degree +12 at t.* If Smith is the man in the blue hat, this might be it. In that case, the intrinsic value of *the man in the blue hat is happy to degree +12* is +12.

However, there is reason to be uncomfortable with this proposal. It conflicts with what we may call the "necessity principle"

> NIV: For any state of affairs, p, and number, n, if IV(p) = n, then it's necessary that IV(p) = n.[14]

Since a given state of affairs might be made true by basics with different intrinsic values at different worlds, the present proposal implies that a given state of affairs might have different intrinsic values at different worlds. It also seems obscure. Consider *someone is happy.* Suppose several different people are in fact happy. Which of these is such that a basic about him "makes true" the generalization that *someone is happy*?[15]

Michael Zimmerman has suggested[16] that the search for a precise number indicating intrinsic value in every case is a wild goose chase. Maybe some things are not intrinsically good and not intrinsically neutral either. For example, disjunctions of good and bad. Maybe some other things, such as "vague and indeterminate intrinsic goods" are intrinsically good, but not to any specific degree. Maybe they are indeterminately intrinsically good. This means that basics bear their intrinsic values in a manner different from non-basics. Every basic has a precise intrinsic value; but some non-basics have indeterminate intrinsic values. It is not clear to me that this is a problem. Suppose we have determinate intrinsic values for all basics. Suppose we have determinate intrinsic values for lives, worlds, and total

---

[14] In Section 16 of *Principia Ethica*, Moore seems to be discussing this principle, though he speaks of the "universal truth" of judgments of intrinsic value. Later in 'The Conception of Intrinsic Value' he again seems to endorse a version of the necessity principle when he says (260-261) that it is *impossible* for one and the same thing to have a certain value at one time or in one set of circumstances, but not at another time or in another set of circumstances. This is complicated by the fact that he seems also to be assuming that nothing can change its intrinsic nature.

[15] A variety of suggestions have been made. Here are three: (a) the intrinsic value of a state of affairs such as someone is happy is equal to the intrinsic value of the least good basic that could make it true; (b) such things have vague intrinsic values; they are good, but they do not have any precise amount of goodness – perhaps the value of such a thing can be represented as a range between worst and best basic that could make it true; (c) such things have, as their intrinsic values, the average intrinsic value of basics that make it true. Some of these have been discussed in the literature. None of them seems particularly plausible to me.

[16] Zimmerman presented the outline of this idea in his 'Evaluatively Incomplete States of Affairs' back in 1983. He discusses it further in his 'Virtual Intrinsic Value and the Principle of Organic Unities' [* pp. 401-413 of this volume].

consequences. Why should we be dismayed when we discover that we do not have precise intrinsic values for certain non-basics?[17]

So my point here is this: on any axiology, each basic must get a precise intrinsic value. But states of affairs that are not basics may be left with vague, indeterminate, or undefined intrinsic values.

3. *Completeness.* Following Harman, let us assume that there are two functions, IV and BIV. IV takes us to the regular intrinsic value of a state of affairs; BIV takes us to the basic intrinsic value.

Where p is a basic intrinsic value state, BIV(p) = IV(p). This is in every case a determinate number. I assume that, for any state of affairs, p, if p is a basic intrinsic value state, then there is some number, n, such that n = the basic intrinsic value of p, or BIV(p). But what shall we say about states of affairs that are not basics? What does BIV do in those cases?

It will be convenient to say that if a state of affairs is not a basic intrinsic value state, then its basic intrinsic value is zero. This is nothing more than the claim that such states of affairs do not have any intrinsic value *in the most basic way.* It is consistent with the idea that such states have plenty of intrinsic value in some derived way.

Given these assumptions, the functions BIV and IV differ in an important respect: BIV is "complete" – for every state of affairs (whether basic or not), BIV takes us to some precise number (of course, the number is often zero). But IV is not "complete" – for many states of affairs, IV is undefined. In the case of many non-basics, p, there is no number, n, such that IV(p) = n.

4. *The Supervenience Thesis.* Each axiology must identify the basics in such a way that all the intrinsic value at any world is determined by the basics true there. In the case of a simple axiology such as our sample eudaimonism, the basics are all things like *Jones is happy to degree +3 at noon today.* Each such thing that's true at a world contributes its intrinsic value to the world. The value of the world is the sum of the values of these true basics. On this axiology, nothing else can affect the value of the world. As a result, any two worlds that are alike with respect to basics will have to be alike with respect to intrinsic value. In other words, the value of a world supervenes upon the values of the basics that are true there.

5. *Indefeasibility.* If we have chosen our basics correctly, then we have chosen our basics in such a way that their value is indefeasible, in a certain specifiable sense. This is a complex matter and a number of possible views have been proposed.[18]

---

[17] For an argument against this idea, see Chisholm, *Brentano and Intrinsic Value*, 81-82 [* p. 313 of this volume].

[18] Thomas Hurka discusses two conceptions of defeat in his very insightful paper, 'Two Kinds of Organic Unity'.

According to one conception of defeat, the intrinsic value of a given state of affairs may vary from circumstance to circumstance. Those who view things this way may start out by supposing, for example, that *Jones being happy to degree +12 at noon today* (or “J12n”) usually has an intrinsic value of +12. But they may go on to suppose that the intrinsic value of J12n is somehow diminished when Jones gets this happiness undeservedly; or when he gets his happiness as a result of seeing that Smith is suffering. They might think that under those circumstances IV(J12n) is less than +12; maybe it is zero. They might also suppose that J12n sometimes has greater intrinsic value, as for example when Jones has been very good and deserves a lot of happiness but so far has not received any.[19]

However, there is another way to view these phenomena of defeat and enhancement. We may be moved by the notion that a thing’s intrinsic value is supposed to be the value it has “in itself”, or “in virtue of its own nature”. We may think, then, that since J12n has remained precisely the same “in itself”, its intrinsic value cannot change no matter what circumstances it occurs in. We therefore think that J12n (as well as everything else) has its intrinsic value of necessity.

If we want to view things in this way, we may say instead that J12n has precisely the same intrinsic value (+12) in every possible circumstance. Yet, when J12n occurs as part of certain larger circumstances, the value of J12n may be “obliterated”. By this I mean to suggest that the “larger” state of affairs somehow fails to reflect the intrinsic value of J12n. Consider, for example, the state of affairs

D: Jones being happy to degree +12 at noon today when Jones does not deserve to be happy at all.

We may suppose that D has an intrinsic value of 0. In this case, D contains a good part (J12n); it contains no bad part, but its intrinsic value is significantly less than the intrinsic value of this good part. J12n continues to have its intrinsic value, but it seems to have been made irrelevant.

When I say that the intrinsic value of J12n is “made irrelevant”, part of what I mean is that when J12n occurs in the context of D, then although J12n continues to have positive intrinsic value, that value does not help to make the world better. Because it has been defeated by its context, the intrinsic value of J12n fails to make any contribution to the intrinsic value of the world.

Let us say that the intrinsic value of a state of affairs, p, is *obliterated* by p’s occurrence in a larger state of affairs, q, iff p is part of q and the intrinsic value of p,q worlds directly reflects the value of q, but not of p.

This second view about intrinsic value gives us yet another way to distinguish between regular, old-fashioned intrinsic value and basic intrinsic value. For we can say that regular intrinsic values can be obliterated, but basic intrinsic values cannot be obliterated. In other words, what happened to the intrinsic value of J12n in the

[19] Chisholm seems to see things in this way in his Presidential Address, ‘The Defeat of Good and Evil’. I inadvertently suggested that I endorse it in my ‘Adjusting Utility for Justice’.

example just discussed could never happen to the basic intrinsic value of a basic intrinsic value state.

Since the basics on any axiology include all the things with non-zero basic intrinsic value, this thesis puts a constraint on what we may identify as the basics. We have to be sure to choose them in such a way that their values will never be obliterated. The point is that if BIV(p) = +n, then any world with p true in it must as a result be n units better than the world just like it except that p is false.[20]

Those who believe that the possibility of obliteration is pervasive may, as a result, be forced into a sort of axiological holism. Suppose we start out thinking that something like J12n has an intrinsic value of +12. But suppose we come to think this value may be obliterated when J12n occurs in the context of D; and suppose we also think that the value of D may be obliterated when D occurs in some wider context. Suppose, more generally, we think that every apparently intrinsically valuable state of affairs runs the risk of having its value obliterated. Whenever we are inclined to assign an intrinsic value to a state of affairs, we stop ourselves and request further information about the circumstances of its occurrence. This fear of obliteration does not relax until we reach whole possible worlds. Then we rest assured that we have found something whose value cannot be obliterated by its occurrence in some larger defeating situation. (We reach this conclusion, of course, entirely because we think that worlds are so large that they cannot occur in any larger situations.)

If we are attracted to this way of viewing intrinsic value, and we agree that basic intrinsic value cannot be obliterated, then we will have to conclude that the only states of affairs that are basics are whole possible worlds. If we reach this conclusion, we will be committed to “world holism”. This would be unfortunate, for it would mean in effect that worlds are organic unities whose basic intrinsic values are not functionally dependent upon the basic intrinsic values of the things that happen within them. But the view seems to me to be coherent, as does a similar view we might call “life holism”.

6. *Familiality; Falling into families.* On the eudaimonistic axiology I have been imagining, each basic is a pure attribution of some amount of happiness to some person at some time. If we hold the person and the time constant, but allow the amount of happiness to vary, we get a certain set of basics. This list suggests such a set:

J+3n: Jones is happy to degree +3 at noon today.
J+2n: Jones is happy to degree +2 at noon today.
J+ln: Jones is happy to degree +1 at noon today.
J0n: Jones is happy to degree 0 at noon today.
J–1n: Jones is happy to degree –1 at noon today.
J–2n: Jones is happy to degree –2 at noon today.

[20] Here, I am speaking loosely. In many cases there is no world that differs only with respect to one selected state of affairs.

J–3n: Jones is happy to degree –3 at noon today.

Consider the set containing these and other states of affairs like them. Each of these is directly about Jones. Each is also directly about noon today. Each is directly about some degree of happiness, but they differ with respect to the precise amount of happiness attributed. Thus, if Jones exists at noon today, exactly one member of this set must be true. Let's use the term "Jones noon family" for the set containing all of these plus the state of affairs *Jones does not exist at noon today* (which I am not counting as a basic intrinsic value state, even though I am assuming that it is a member of the family). Now we can say: exactly one member of the Jones noon family must be true.[21]

Since the basics on each axiology are pure attributions of whatever is asserted by that axiology to be the fundamentally intrinsically good property, it follows that the basics selected by any axiology will have to fall into families such as this one. In general, suppose some axiology selects some relation, *x,n,t|Rx,n,t*, as the source of all intrinsic value; suppose this relation relates a person, x, to a number, n, and a time, t. Then, for any selected pair of person, S, and time, t, there will be an S,t-family. The members of that family will be all the basics about S and t, differing only in the precise amount of R that they attribute to S at t (plus *S does not exist at t*). Exactly one member of this family must be true.

We can then make a general statement about basic intrinsic value states: they fall into families. In some cases, admittedly, the families might be fairly degenerate.

Facts about families may shed some light on the concept of "isolation". In a number of places, Moore claimed that the intrinsic value of a thing is the amount of value it would have "in isolation".[22] Moore's remark is suggestive, but problematic. It becomes even more problematic if we assume (as I do) that the fundamental bearers of intrinsic value are states of affairs. What could be meant by saying that some state of affairs exists "in isolation"?[23]

Many traditional axiologies yield families that are logically independent. In the case of any such axiology, if p is a member of one family, and q is a member of another family, then p and q are logically independent. There are worlds where both are true, worlds with neither true, and yet others with one but not the other.[24]

---

[21] I first encountered the idea that basic intrinsic value states fall into families in Warren Quinn's, 'Theories of Intrinsic Value'. Quite a bit of what I say in this paper is influenced by Quinn.

[22] Moore, *Principia Ethica*, 91.

[23] For a perceptive discussion of problems about the isolation test, see Eva Bodanszky and Earl Conee, 'Isolating Intrinsic Value' [* pp. 11-13 of this volume].

[24] On more complex axiologies, this may sometimes fail. Sometimes a member of one family will stand in a logical relation to a member of another. Thus, for example, we can consider an axiology according to which pleasure and knowledge are both intrinsically good. Then consider this pair:

J10: Jones is pleased to degree +10 at t.
SK10: Smith knows that Jones is pleased to degree +10 at t.

This fact about simple axiologies suggests a possible interpretation of Moore's isolation thesis. Let us confine our attention to axiologies according to which every basic is such that there is some possible world in which it is the sole true basic. On these axiologies, if p is a basic and true at w, then no other member of its family is true at w. Furthermore, (in the case of simple axiologies) each basic is such that it's possible for it to be true even though no extrafamilial basic is true. Therefore, in these cases, there is a world at which no other basic is true.

Let us say that if p is a basic, and w is a world at which p is the only true basic, then p is *basically isolated* at w. I have argued that for many basics, there is a world at which it is basically isolated. A restricted version of the isolation thesis may now be stated: if p is a basic, and there is a world at which it is basically isolated, then IV(p) is equal to the intrinsic value of any world at which p is basically isolated.

Of course this won't work for axiologies that permit basics to entail basics in other families.

4. SOLUTIONS TO PUZZLES

At the outset I mentioned some puzzles and I claimed that the appeal to basics might be useful for their solution. Let us briefly review the proposed solutions.

i. The first puzzle concerned a contradiction into which we may fall when we try to state an axiological theory. If we try to formulate the simplest sort of hedonism, we may be inclined to say, for example, that (i) hedonic (doloric) states are the only intrinsic goods (bads); and (ii) a world or other complex thing is intrinsically good (bad) iff it contains more (less) pleasure than pain. As I pointed out in Section 1, this is a contradiction.

I propose that we make use of basics in our statement of any axiological theory. Thus, for example, if we want to formulate a simple form of eudaimonism, we should start out by identifying the things that are basic intrinsic value states on that axiology. We might say that a *basic eudaimonistic state* is any pure attribution of happiness, something like:

J12n: Jones is happy to degree +12 at noon today.

And then we could say that on our eudaimonistic axiology, the basic intrinsic value states are all and only these basic eudaimonistic states. We could go on to say that the intrinsic value of each such thing is equal to the precise amount of happiness attributed. Thus, IV(J12n) = +12.

We might want to say that each of these is intrinsically good; they are members of distinct families, and SK10 entails J10. Thomas Hurka discusses this sort of axiology in detail in his 'Virtue as Loving the Good' as well as in his *Virtue, Vice, and Value.*

Then we could say that the intrinsic value of a world, life, or total consequence is equal to the sum of the intrinsic values of the basics that are true within it.[25] Thus, on this axiology, J12n will inevitably contribute exactly +12 units of intrinsic value to any world at which it occurs. Given this way of calculating the intrinsic value of a world, we can conclude by saying that a world, life, or total consequence is intrinsically good (bad) if its intrinsic value is greater (less) than zero. If we proceed in this way, our formulation will not be internally inconsistent.[26]

Suppose we prefer a much more complex axiology, such as the one Moore presented in the final chapter of *Principia Ethica.* Moore did not endorse anything simple, like hedonism or eudaimonism. Rather, he said that some of the "great intrinsic goods" are things such as "the love of beauty" and "the hatred of evil". Moore's discussion makes it clear that each of these is in fact a very complicated type of state of affairs involving "proper appreciation", knowledge, and the actual existence of the appreciated object. What then are the basics?

As I see it, one part of Moore's view is that something such as this is a basic intrinsic value state:

> LB: Jones taking aesthetic pleasure of intensity +10 in the beautiful qualities, Fl, F2, F3, etc. of object A at noon today while knowing that object A in fact exists and has F1, F2, F3, etc., and that these are beautiful qualities.

LB is an instance of "the love of beauty". It is intended to be a pure attribution of a complex relation: *x,n,K,y,t|x taking aesthetic pleasure of intensity n in the beautiful qualities in set K of object y at time t while knowing that y in fact exists and has the members of K and that the members of K are beautiful qualities.* The intrinsic value of LB might be thought to depend upon the intensity of the aesthetic pleasure taken. If we wanted to capture some other thoughts Moore suggests, we could consider a slightly more complex variant in which there are places for numbers indicating the extent to which the object has the beautiful properties, and perhaps even more.

Note that LB has many "proper parts", states of affairs that it entails but which do not entail it. Some examples are:

> JP: Jones taking pleasure of intensity +10.
> JKA: Jones knowing that object A exists.

None of these is a basic intrinsic value state on the Moorean axiology. None of these has any basic intrinsic value and it is not clear that any of them has any determinate

[25] Note that we do not say that the value of a world depends upon what basics *exist* within it; that would yield an incorrect number. Since basics are abstract entities, each of them exists at each world. Rather, we say that it depends upon what basics are *true* within it.

[26] Although it may have other problems involving the need to find the sum of an infinite number of addends. It may also have some problems due to the fact that the substance of the axiology is somewhat naive.

(regular) intrinsic value. Similar comments apply to all the other proper parts of LB. Thus, the intrinsic value of LB is not determined by performing some mathematical operation such as addition on the intrinsic values of its parts. When those basically worthless parts come together, something of great value emerges. Perhaps this gives some sense to Moore's insistence that the great intrinsic goods are all organic unities.[27]

As an example of a case of "hatred of evil", we could consider:

HE: Jones experiencing hatred of intensity +10 in the vicious qualities, V1, V2, V3, etc. of person Z at noon today while knowing that person Z in fact exists and has V1, V2, V3, etc., and that these are vicious qualities.

To complete the statement of a Moorean axiology, we would have to identify all the main sorts of basic intrinsic value states, and we would have to provide principles specifying, for each type, the basic intrinsic value of states of that type. Then we could say that the intrinsic value of a life, a total consequence, or a world is a simple function of the intrinsic values of the basics true therein. Again, if this were done carefully, it would be internally consistent and true to the spirit of Moorean axiology.

Thomas Hurka has recently discussed a type of axiological theory that involves a sort of "embedding", or recursion. On one theory of that sort, we could start by saying that pure attributions of pleasure are intrinsically good basics. Then we could say that knowledge of something intrinsically good is also intrinsically good. Now consider this sequence:

1. S1 feeling pleasure of intensity +10 at noon.
2. S2 knowing that S1 feels pleasure of intensity +10 at noon.
3. S3 knowing that S2 knows that S1 feels pleasure of intensity +10 at noon.
4. S3 knowing that S1 feels pleasure of intensity +10 at noon.

On the imagined axiology, each of these would be a basic intrinsic value state and each would independently contribute some intrinsic value to any world in which it is true. Assuming that each has a basic intrinsic value of +10, we might want to say that if these four are the only basics true at a world, then the value of that world is +40.

Thus, though (3) "guarantees" forty units of intrinsic value, I have assigned it a basic intrinsic value of just +10. (3)'s guarantee is of course fulfilled; if (3) is true, then so are (1) and (2) and (4). Since each of these others is also a basic on this axiology, each of them will contribute ten more points to the value of the world.

---

[27] Í do not offer this remark as a general characterization of organic unities.

I prefer to give (3) a basic value of just +10 rather than +40. One reason for this is that I want to appeal to (3)'s basic intrinsic value when I calculate the value of the life of S3. As I see it, anyone endorsing this axiology would want to say that the value of S3's life should be increased by just 10 points as a result of (3)'s truth. Even though the truth of (3) guarantees 40 units of value, to increase the value of S3's life by 40 points would be excessive. S3's life is only 10 points better as a result of the truth of (3). So I say that BIV(3) = +10. Then the value of the world comes out right, as does the value of each of the lives.

ii. The second puzzle concerns the concept of instrumental value. Clearly, we cannot say that the instrumental value of a state of affairs is equal to the sum of the regular intrinsic values of all the things it causes. To do so would invite double counting in cases such as the case involving Bob who reads a delightful book. Recall this state of affairs:

B: Bob reads at 9:00PM on Monday evening.

I imagined a case in which B makes Bob happy to degree +10 while he reads, and sits in his easy chair, and wears his eyeglasses. In this case, B has many consequences, including the following:

C1: Bob being happy to degree +10 at 9:00PM on Monday evening.
C2: Bob being happy while reading.
C3: Someone being happy to degree +10.
C4: The man in Bob's easy chair being happy to degree +10.
C5: Bob being happy at least to degree +9.
C6: Bob being happy to degree +10 while sitting in his easy chair and wearing his eyeglasses at 9:00PM on Monday evening.
C7: Either Bob being happy to degree +10 or Babe being happy to degree +15.

Since each of these is a consequence of B, and each seems to have some positive (regular) intrinsic value, the simple aggregative principle yields an incorrectly high estimate of the instrumental value of B.

As a first step toward a solution, I propose:

EV3: The instrumental value of a state of affairs, p, at a world, w, is the sum of the intrinsic values of all the basic intrinsic value states, q, such that p causes q at w.

Again assuming a simple sort of eudaimonism, EV3 implies (much more plausibly) that the instrumental value of B is +10.

Some philosophers think that a state of affairs gets a boost in instrumental value if it *prevents evil.* Thus, we might say that it's instrumentally good to get your

cavities filled since this prevents the unhappiness you would experience if you left them unfilled. EV3 makes no provision for such preventive instrumental value. It is easy enough to incorporate this idea. We do it in two steps as follows:

First we introduce the notion of "preventive value". To find the preventive value of a state of affairs, p, first consider all the states of affairs it prevents; then locate the basic intrinsic value states among these; then find the sum of their intrinsic values. Since it's good to prevent evil, but bad to prevent good, the preventive value is equal in amount but opposite in sign to the sum of the intrinsic values of the basics prevented. So if some state of affairs prevents basics whose intrinsic values sum to –10, then that state of affairs has a preventive value of +10.

Now we can define instrumental value:

> EV4: The instrumental value of a state of affairs, p, at a world, w, is the sum of the intrinsic values of all the basics, q, such that p causes q at w plus the preventive value of p at w.[28]

iii. Our third puzzle concerned the calculation of the intrinsic values of complex things such as lives, total consequences, and possible worlds. Since I have incorporated aggregative principles into the formulation of the sample axiologies, it is not necessary to devote much attention to this puzzle. The solution is the same in all cases: the intrinsic value of a complex thing is in each of these cases equal to the sum of the intrinsic values of the basic intrinsic value states contained within.

I acknowledge that this leaves the intrinsic values of certain states of affairs undefined. For example, consider disjunctions such as:

> Jv: Jones being happy to degree +12 at t1 or Jones being happy to degree +10 at t2.

Even on the simplest sort of eudaimonism, Jv is neither a basic nor a world nor a life. Nothing I have said implies anything about its intrinsic value. So far as I can see, this is not a problem. Let its intrinsic value be undefined. If it is true and a part of the life of Jones at some world, then either its first disjunct or its second disjunct or both will also be included in Jones' life at that world. One or both of those basics will therefore contribute its full value to the value of the life and the world. Thus, the axiological significance of the disjunction is entirely accounted for by the basic intrinsic value states that make it true.

ACKNOWLEDGEMENT

I have benefited enormously from the generous critical comments of many teachers, students, friends and colleagues. Among these are Eva Bodanszky, Ben Bradley,

---

[28] I have benefited from discussions on this topic with Ben Bradley. See his doctoral dissertation, *Species of Goodness*, University of Massachusetts, Fall, 1998.

Phil Bricker, Roderick Chisholm, Earl Conee, Ish Haji, Shelly Kagan, Owen McLeod, Ned Markosian, Ted Sider and Michael Zimmerman. Some of the ideas are developed from written work (cited below in the references) of Roderick Chisholm, Gilbert Harman, Thomas Hurka, and Warren Quinn.

I acknowledge that I have not incorporated every suggestion that I have received. Nevertheless, I am grateful for them all.

I have presented various bits of this paper in seminars and talks over a period of many years. Many participants in those discussions provided useful insight and challenge. Although I cannot now cite them individually, I am grateful to them all.

REFERENCES

Aristotle, *Nicomachean Ethics.*

Bodanszky, E. and Conee, E. (1981), 'Isolating Intrinsic Value', *Analysis* 41, 51-53 [* pp. 11-13 of this volume].

Bradley, B. (1998), *Species of Goodness*, University of Massachusetts.

Brandt, R. (1967), 'Hedonism', in P. Edwards (ed.), *The Encyclopedia of Philosophy*, vol. 4, New York: Macmillan Publishing Co. & The Free Press, pp. 432-435.

Chisholm, R. (1986), *Brentano and Intrinsic Value*, Cambridge: Cambridge University Press.

Chisholm, R. (1968), 'The Defeat of Good and Evil', *Proceedings and Addresses of the American Philosophical Association* Volume XLII – 1968-69, pp. 21-38.

Chisholm, R. (1975), 'The Intrinsic Value in Disjunctive States of Affairs', *Noûs* 9, 295-308 [* pp. 229-39 of this volume].

Feldman, F. (1978), *Introductory Ethics*, Englewood Cliffs: Prentice-Hall, Inc.

Feldman, F. (1986), *Doing the Best We Can: An Essay in Informal Deontic Logic*, Dordrecht: Reidel.

Feldman, F. (1995), 'Adjusting Utility for Justice', *Philosophy and Phenomenological Research* 55, 567-585.

Harman, G. (1967), 'Toward a Theory of Intrinsic Value', *Journal of Philosophy* 64, 792-804 [* pp. 349-60 of this volume].

Hurka, T. (1992), 'Virtue as Loving the Good', in E. F. Paul et al. (ed.), *The Good Life and the Human Good*, Cambridge: Cambridge University Press, pp. 149-168.

Hurka, T. (1998), 'Two Kinds of Organic Unity', *Journal of Ethics* 2, 299-320.

Hurka, T. (2000), *Virtue, Vice, and Value*, New York: Oxford University Press.

Lewis, C.I. (1946), *An Analysis of Knowledge and Valuation*, LaSalle, Illinois: Open Court Publishing Co.

Moore, G.E. (1993), *Principia Ethica*, revised edition edited and with an introduction by Thomas Baldwin, Cambridge: Cambridge University Press.

Moore, G.E. (1922), 'The Conception of Intrinsic Value', in *Philosophical Studies*, London: Routledge & Kegan Paul, pp. 253-275.

Plato, *Republic.*

Quinn, W. (1974), 'Theories of Intrinsic Value', *American Philosophical Quarterly* 11(2), 123-132.

Zimmerman, M.J. (1983), 'Evaluatively Incomplete States of Affairs', *Philosophical Studies* 43, 211-224.

Zimmerman, M.J. (1999), 'Virtual Intrinsic Value and the Principle of Organic Unities', *Philosophy and Phenomenological Research* 59, 653-666 [* pp. 401-413 of this volume].

CHAPTER 32

M. J. ZIMMERMAN

VIRTUAL INTRINSIC VALUE AND THE PRINCIPLE OF ORGANIC UNITIES

1. INTRODUCTION

In 1903, G. E. Moore introduced us to the principle of organic unities. In his words, this principle says: "The value of a whole must not be assumed to be the same as the sum of the values of its parts."[1] He was talking, of course, of intrinsic value. Since that time the principle has been frequently discussed, and it has usually been endorsed by those who have discussed it.[2]

Purported examples of organic unities abound. Prominent among these are apparently anomalous cases of pleasure and displeasure. Whereas pleasure has seemed to many to be intrinsically good, pleasure taken in something bad has not; and whereas displeasure itself has seemed intrinsically bad, displeasure taken in something bad has not.[3] Note that such cases apparently cannot be construed simply as cases where the value of one part of the whole situation is counterbalanced by the value of another part. When malicious pleasure has as its object something that in fact does not exist but which the person in question believes to exist, this is almost as troubling, ethically, as when the object does exist. Yet in such a case we cannot appeal to the badness of the object as a counterbalance to the goodness of the pleasure.[4] And, in the case of displeasure taken in something bad, there is no apparent candidate for a counterbalance to the badness of the displeasure, whether or not its object exists. Other anomalous cases involving pleasure and displeasure are those where the pleasure is undeserved and the displeasure deserved.[5] Still other cases of alleged organic unities concern the order

[1] Moore (1903), p. 28, italics deleted.
[2] See especially, in recent years, Chisholm (1986) and Lemos (1994).
[3] See Brentano (1969), pp. 90-91 (this book was first published in the original German in 1889); Moore (1903), pp. 208 ff.; Chisholm (1986), pp. 76 ff. [* pp. 309 ff. of this volume]; Lemos (1994), pp. 35 ff.
[4] See Chisholm (1986), p. 76 [* p. 309 of this volume].
[5] See Ross (1930), p. 72; Chisholm (1968-69), p. 30.

T. Rønnow-Rasmussen and M. J. Zimmerman (eds.), Recent Work on Intrinsic Value, 401-413.
© 2005 *Springer. Printed in the Netherlands.*

in which things occur, such as whether there is a progression from bad to good or from good to bad, whether goods are varied or occur in homogeneous groups, and so on.[6]

I think that the principle of organic unities is false, and in this paper I shall try to explain why. (More accurately: I think that the principle needs reformulation and that, when properly formulated, it should be declared false. The reason for this qualification will become apparent later.) One way to argue that the principle is false is to claim that those who diagnose certain situations as involving organic unities have overlooked relevant value-laden parts or aspects of the situation. For example, it is clear that our inclination to frown on pleasure taken in something bad is rooted in our assessment of the emotion as being somehow inappropriate to its object. Why, then, not say that this inappropriateness is a part or aspect of the overall situation that bears its own intrinsic value – a negative one – which counterbalances the goodness of the pleasure itself?[7] I reject this approach. Even if a suitable account can be given of how the inappropriateness of the emotion is a bad part of the overall situation, still the appeal to counterbalancing seems to me in this case to be mistaken. For the badness of the inappropriateness (if it is bad) does not outweigh the goodness of the pleasure, and this is because (contrary to what this critic of the principle of organic unities says, and also to what most proponents of the principle say) the pleasure simply is not good.

Or so I shall argue. But I shall not attempt to do this by arguing that the pleasure in such a case is bad.[8] For the pleasure is not bad, either. Nor is it neutral. To think that it must be one or the other of these is to overlook the fact that intrinsic value can be merely virtual rather than actual.

2. THE BEARERS OF INTRINSIC VALUE

Whether or not the principle of organic unities is true depends of course on what sort of thing may be said to bear intrinsic value and to be such that it is either a whole or part of some whole. Although Moore himself never settled on an answer to this question, many answers have been proposed by others, most notably these: states of affairs,[9] properties,[10] and facts.[11] For reasons which space does not permit me to discuss here, I think we should say that *states of individuals* are the bearers of intrinsic value, where these are understood as *concrete occurrences* consisting in individuals exemplifying certain properties. Thus, and for example, when it is said that pleasure itself is intrinsically good, this would be understood as the claim that any and every episode of pleasure – any and every concrete occurrence that consists in someone's or something's

[6] See Chisholm (1986), pp. 70 ff. [* pp. 306 ff. of this volume]; Lemos (1994), pp.37 ff., 199-200 [* pp. 322-23 of this volume].

[7] Perhaps Brentano has this in mind at Brentano (1969), p. 23, n. 1.

[8] This is what is apparently implied in Dancy (1993), pp. 61 ff. Yet Dancy declares pleasure in other cases to be good. This seems to me inconsistent with how the notion of *intrinsic* value is to be understood, but the issue is too complex to address here.

[9] See Chisholm (1986), pp. 60 ff.

[10] See Butchvarov (1989), p. 14.

[11] See Ross (1930), p. 137; Lemos (1994), pp. 22-23 [* pp. 181 ff. of this volume].

being pleased – is intrinsically good. The property of being pleased is of course "doing the work" here, in that it is this that makes episodes of pleasure intrinsically good. Nonetheless, it is the episodes, and not the property, that bear this value. (Where the property is not exemplified, there is – to that extent – nothing intrinsically good, even though the property itself exists.)

States are to be individuated finely. Suppose that on some occasion Peter is undeservedly pleased. Do "Peter is pleased" and "Peter is undeservedly pleased" refer to one state or two? The answer is: two. (This stems from the fact that being pleased and being undeservedly pleased are themselves distinct properties.) This answer is in keeping with what both proponents and opponents of the principle of organic unities typically want to say. Proponents will typically say that Peter's being pleased is a good thing but his being undeservedly pleased is a bad thing. If there were just one state here and not two, how could this be so? One would have to resort to some such locution as "this state is intrinsically good under one description but not under another," and this seems incoherent; for how could something's *intrinsic* value be relative to some contingent description of it? And opponents such as myself will likewise wish to distinguish between the intrinsic value to be attributed to such states as Peter's being undeservedly pleased and that to be attributed to such states as Peter's being pleased.

It is natural to say that a state such as Peter's being pleased is a *part* of the state consisting in his being undeservedly pleased. Perhaps this can be accounted for as follows: the existence of the former state is *entailed* by the existence of the latter.[12] On this account, Peter's being pleased will also be a part of Peter's being pleased and Paul's being displeased, which again seems a natural thing to say. Perhaps it is not so natural to call this conjunction of states *itself* a state, but for convenience I shall so call it.

Finally, I shall assume that it is *only* states (including conjunctions of states) that have and can have intrinsic value. In saying this, I am agreeing at least partially with such proponents of the principle of organic unities as Roderick Chisholm and Noah Lemos, who believe that ascriptions of intrinsic value can all be made with respect to a single type of entity. But these authors appear implicitly to rely, in their discussion of this principle, on the further claim that *all* entities of the relevant type have intrinsic value. It is this claim that I think is false and whose falsity allows for a rejection of the principle. More precisely, I think it is false to say that all states have *actual* intrinsic value.

3. INTRINSIC VALUE AND FAVOR

Many philosophers have noted a connection between a state's having a certain intrinsic value and the fittingness of a certain attitude towards that state. There are variations among the proposals that have been offered, but the main gist seems to be this:

---

[12] This corresponds to suggestions made in Chisholm (1986), p. 73 [* p. 308 of this volume], and Lemos (1994), pp. 33-34. Just what constitutes the sort of entailment in question is a difficult issue that I shall not try to resolve here. There is reason to think it must be more than simply strict implication.

(I) Necessarily, state S is intrinsically good [bad] if and only if the contemplation of S and S alone requires that one favor [disfavor] S.[13]

I have called the pro-attitude in question "favor" and the corresponding anti-attitude "disfavor," rather than use the more usual terms "love" and "hate,"[14] for the reason that the latter terms have too many irrelevant connotations and, moreover, suggest considerable intensity of feeling, whereas the former terms, being less common, are better suited to function as terms of art in this context.

I think that (I) should be accepted, and I will build my case against the principle of organic unities on the basis of it. I shall not seek to argue for (I) in turn but will have to rest content with the following observations.

First, when it is said that the contemplation of "S and S alone" requires a certain attitude, this is intended to indicate that it is just S and nothing else (in particular, not S's circumstances or S's consequences) that is the object of contemplation. It is this that will determine whether S is *intrinsically* worthy of favor or disfavor.

Secondly, the term "require" is borrowed from Chisholm, who takes it as primitive.[15] In this he is followed by Lemos, who however qualifies the concept at issue as that of *ethical* requirement.[16] This seems to me quite correct. Chisholm does not distinguish between ethical and non-ethical requirement. As illustrations of one thing requiring another, he gives the following:

> ...promise-making requires – or calls for – promise-keeping; being virtuous requires...being rewarded; the dominant seventh requires the chord of the tonic; one color in the lower left calls for a complementary color in the upper right.[17]

But while the first two illustrations concern ethics, the last two concern aesthetics. It may be that promise-making requires promise-keeping, but, if it does, this is an ethical matter. If one color calls for another that is complementary rather than incongruous, that is a matter of aesthetics. (One indication of what is at stake here is this. We would say that, if someone who contemplates something intrinsically good [bad] displays the required favor [disfavor], this is evidence of that person's *ethical* sensitivity.)

Thirdly, contrary to the assertions of some,[18] it seems clear that favor and disfavor (however precisely these attitudes are to be understood, and I shall offer no account of them here) come in degrees. This permits us to refine (I) as follows:

---

[13] Cf. Brentano (1969), p. 18; Ross (1939), p. 279; Ewing (1948), Ch. 5; Chisholm (1986), Ch. 5; Lemos (1994), Ch. 1 [* pp. 17-31 of this volume]; and Goldstein (forthcoming).

[14] See Brentano (1969), pp. 18, 137 ff.; Chisholm (1986), Chs. 3, 5; Lemos (1994), pp. 6 ff. [* pp. 20 ff. of this volume]. Lemos makes mention of "favor" and "disfavor" and employs them on occasion. Blanshard also uses these terms in Blanshard (1961), pp. 280 ff.

[15] Chisholm (1964) and (1986), p. 52.

[16] Lemos (1994), p. 12 [* p. 24 of this volume].

[17] Chisholm (1964), p. 147.

[18] See Brentano (1969), pp. 25-26; Chisholm (1986), pp. 19-20; Lemos (1994), p. 12 [* p. 24 of this volume].

(I') Necessarily, S is intrinsically good [bad] to a certain degree if and only if the contemplation of S and S alone requires that one favor [disfavor] S to a corresponding degree.

Thus (to use numbers in a purely *ad hoc* fashion): what is intrinsically good to degree 10 requires favor to a corresponding degree 10; what is intrinsically bad to degree 5 requires disfavor to a corresponding degree 5; and so on.

Fourthly, favor and disfavor are here to be construed as "overall" attitudes, attitudes that one takes "on balance" towards states. (This is analogous to giving a movie a vote of "thumbs up" or "thumbs down." While one might like the movie in certain respects and not in others, one's overall, on-balance assessment cannot be a vote of "*both* thumbs up *and* thumbs down.") Thus, if one favors a certain state, one doesn't disfavor it. But the reverse does not hold; for indifference is possible. One is indifferent (in the relevant sense) towards some state if and only if one neither favors nor disfavors it. (This is like a vote of "*neither* thumbs up *nor* thumbs down." Actually, this account of indifference is too rough; important qualification will be provided in the next section.) Given this notion of indifference, we can add to the foregoing account of intrinsic goodness and badness the following account of intrinsic neutrality:

(II) Necessarily, S is intrinsically neutral if and only if the contemplation of S and S alone requires that one be indifferent regarding S.[19]

4. THE ISOLATION METHOD

Moore says:

> The method which must be employed in order to decide the question 'What things have intrinsic value, and in what degrees?' has already been explained... In order to arrive at a correct decision on the first part of the question, it is necessary to consider what things are such that, if they existed *by themselves*, in absolute isolation, we should yet judge their existence to be good; and, in order to decide upon the relative *degrees* of value of different things, we must similarly consider what comparative value seems to attach to the isolated existence of each.[20]

Lemos calls this "ontological isolationism" and quite rightly says that it is unworkable; for every state is such that it cannot exist in isolation, since its existence implies the existence of countless other things.[21] Lemos contrasts Moore's approach with that employed by Chisholm and exemplified in (I), (I'), and (II) above; he calls this

[19] On pp. 13-14 of Lemos (1994) [* pp. 25-27 of this volume] an interesting distinction is drawn between four types of neutrality. For reasons that space does not permit me to develop, I decline to follow Lemos's lead here.
[20] Moore (1903), p. 187. Cf. Moore (1912), p. 102, and Ross (1930), pp. 68-69, 134.
[21] Lemos (1994), pp. 10-11 [* p. 23 of this volume].

"intentional isolationism." Even if no state can exist by itself, it can be contemplated in isolation, in that it and it alone can constitute the object of contemplation.

Moore employs the isolation method in order to argue against the hedonistic view that pleasure and pain are the only things of non-neutral intrinsic value.[22] Others have followed suit.[23] But most of these authors have nonetheless agreed with the hedonistic view that pleasure is intrinsically good (and pain intrinsically bad), even if such goodness (and badness) may be defeated within the context of a "wider" organic unity.[24] This seems to me a mistake.

Consider ontological isolationism first. Since pleasure cannot exist all by itself, the best we can do here is to imagine a world as "bare" as possible apart from the fact that in it someone is pleased. I take it that this world would involve the pleasure's being neither appropriate nor inappropriate, neither deserved nor undeserved, since these would be further, ethically significant facts about the pleasure whose admission would be inconsistent with this "bare-bones" approach.[25] But then Moore's claim that there is an organic unity when pleasure is taken in something bad cannot be sustained;[26] for here we do *not* have an instance of that pleasure which has been independently identified as good constituting part of a whole that is not good. The pleasure independently identified as good is not an inappropriate pleasure; the pleasure that constitutes part of the whole in question is an inappropriate pleasure.

This sort of problem is obviated by intentional isolationism. Here it can be said that the pleasure that is independently identified by this method is the very same pleasure that constitutes part of the wider situation when pleasure is taken in something bad. And here the claim is that contemplation of the pleasure as such requires that one favor it. But why accept this claim? Once it is recognized that, whenever pleasure occurs, it must be either appropriate or inappropriate or neither, either deserved or undeserved or neither, it seems to me that, if appropriate or deserved pleasure is intrinsically good but inappropriate or undeserved pleasure is intrinsically bad, then what contemplation of pleasure *as such* requires is that one *withhold* either favor or disfavor or indifference. One should *refrain* from adopting any attitude towards it until all further relevant information (concerning appropriateness and desert and other matters, too, if they are relevant) has been supplied. Then and only then – once the "whole" situation has been presented for contemplation – should one adopt an attitude of favor, disfavor, or indifference towards *it.* If appropriate or deserved pleasure is intrinsically good but inappropriate or undeserved pleasure is intrinsically bad, then, I would say, pleasure as such is neither intrinsically good, nor intrinsically bad, nor intrinsically neutral.

Notice that I have talked here of indifference as an attitude, as favor and disfavor are attitudes. Indeed, I understand it to be an overall attitude, just as favor and disfavor are. In the last section I said that one is indifferent regarding some state if and only if one

[22] Moore (1903), Ch. 6; (1912), Ch. 7.

[23] E.g., Ross (1930), Ch. 5; Chisholm (1986), p. 60; Lemos (1994), Ch. 6.

[24] Moore (1903), pp. 87 ff.; Chisholm (1986), pp. 76 ff. [* pp. 309 ff. of this volume]; Lemos (1994), pp. 36 ff.

[25] Here "inappropriate" and "undeserved" mean something stronger than just "not appropriate" and "not deserved."

[26] Moore (1903), pp. 208-9.

neither favors nor disfavors it. But now we can see that this isn't quite right; while the "only if" holds, the "if" does not. For it can happen that one neither favors nor disfavors some state because one lacks *any* overall attitude towards it, including that of indifference. Consider the movie analogy again. Suppose that one has not yet seen the movie; then one is not yet in a position to give it *any* evaluation – not a "thumbs up" (which is analogous to favor), nor a "thumbs down" (analogous to disfavor), nor even a "neither thumbs up nor thumbs down" (analogous to indifference). Thus indifference involves not merely the absence of favor and disfavor but the presence of a third attitude, and it is in this way that it can happen that one displays, indeed that one is *required* to display, *none* of the attitudes of favor, disfavor, and indifference.

I believe that what has just been said concerning pleasure can be applied to all cases where organic unities have traditionally been diagnosed. In all such cases, the "parts" of the "wholes" to which an intrinsic value has been ascribed are such that one ought in fact to withhold any overall attitude of favor, disfavor, or indifference towards them. Consider what is surely an oversimplified case. Suppose that the correct view were that a good person's experiencing pleasure is intrinsically good (regardless of any question of appropriateness), that a neither good nor bad person's experiencing pleasure is intrinsically neutral, and that a bad person's experiencing pleasure is intrinsically bad. What, on this view, should we take the intrinsic value of pleasure itself to be? My answer is: it has no such value. Consider a world in which nothing of intrinsic value occurs except that which is involved in a bad person's experiencing pleasure. I would say that, on the view in question, *nothing* good occurs; there is *nothing* such that the contemplation of it requires that one favor it.

Even if this is correct, though, and the pleasure in this case has no *actual* intrinsic value, it may be said to have a *virtual* intrinsic value. Understanding (I') and (II) to be statements that concern *actual* intrinsic value, we can now supplement them with the following:

(III) Necessarily, S is virtually intrinsically good [bad] to a certain degree if and only if, for some state S' whose existence entails that of S, the contemplation of S' and S' alone requires that one favor [disfavor] S' to a corresponding degree;

(IV) Necessarily, S is virtually intrinsically neutral if and only if, for some state S' whose existence entails that of S, the contemplation of S' and S' alone requires that one be indifferent regarding S'.

In this case, then, the pleasure in question may be said to be virtually intrinsically bad. (In other cases, of course, a pleasure may be virtually good or neutral rather than bad. Thus virtual intrinsic value can vary in a way in which actual intrinsic value cannot. Wherever the exemplification of a certain property has a certain actual intrinsic value, any other exemplification of that property must have the same intrinsic value.) We can also say that a certain state *merely* has a certain virtual intrinsic value when it has that

value but lacks any actual intrinsic value. On the view under discussion, any pleasure is as such *merely* virtually intrinsically valuable.

5. CONDITIONAL INTRINSIC VALUE

It is apparent that I am taking actual intrinsic value to be in a certain sense unconditional, while virtual intrinsic value is conditional. In so doing, I am echoing the suggestions of others who have not been entirely won over by what Moore has had to say about organic unities. For example, W. D. Ross at times talks of pleasure as not being good in itself but rather as being *prima facie* good in itself.[27] H. J. Paton has said something akin to this,[28] and Chisholm has reported that both Franz Brentano and St. Thomas Aquinas at times made remarks in a similar vein.[29]

I think that these remarks are all on the right track. But I hesitate to use either the term “conditional” or the term “prima facie” in this context because, associated as they so often are with the issues of conditional and prima facie obligation, they are liable to mislead. For example, Ross says:

> Pleasure seems, indeed, to have a property analogous to that which we have previously recognized under the name of conditional or *prima facie* rightness. An act of promise-keeping has the property, not necessarily of being right but of being something that is right if the act has no other morally significant characteristic... And similarly a state of pleasure has the property, not necessarily of being good, but of being something that is good if the state has no other characteristic that prevents it from being good.[30]

This is suggestive but doesn’t seem quite right, for it places undue emphasis on the possibility of pleasure’s being virtually good and neglects the possibility that it be virtually neutral or bad. One indication of the disanalogy between prima facie rightness or obligatoriness on the one hand and virtual intrinsic value on the other is this. When an act is prima facie but not overall morally obligatory, it still has a certain moral value such that its nonperformance is cause for regret.[31] But, I would urge, it is emphatically not the case that, when a certain malicious pleasure has been avoided, there is any cause for regret.[32]

6. OBJECTIONS TO THE ACCOUNT

Several objections may be raised against the account that I have given of virtual intrinsic value. I shall consider four.

[27] Ross (1930), p. 138.
[28] Paton (1942), p. 122.
[29] Chisholm (1986), pp. 96-97, including n. 6.
[30] Ross (1930), p. 138.
[31] Ross (1930), p. 28.
[32] Ironically, Chisholm is in agreement with this. See Chisholm (1986), p. 79 [* pp. 311-12 of this volume].

First, it might be objected that, even on my account, the principle of organic unities is true. Suppose that S' is actually intrinsically good but that no other state, whose existence is entailed by that of S', is actually intrinsically good; that is, any such state is merely virtually intrinsically good. Then the intrinsic value of S' is not the same as the sum of the intrinsic values of its parts. Therefore the principle is true.

In response, let me remind you that I said that the principle of organic unities needs reformulation. It must be acknowledged that, as formulated by Moore and as quoted at the outset of this paper, the principle is indeed shown to be true by a case of the sort just given; but this is trivial. Moore presumably formulated the principle as he did because he thought that all parts of all wholes have actual intrinsic value. On this assumption, his formulation of the principle is sufficient to capture what he took to be his main insight, namely, that the intrinsic values of parts can be not just counterbalanced but *defeated* by the intrinsic value of a whole. It is this idea that I am opposing in this paper. Given the possibility of merely virtual intrinsic value, however, the principle of organic unities needs to be reformulated in order to capture this idea of defeat. I propose the following formulation (for the time being – but see the final section):

> The actual intrinsic value of a state must not be assumed to be the same as the sum of the actual intrinsic values of those states that are a part of it and have such value.

My purpose has been to cast doubt on the existence of organic unities so understood.

A second objection is that I have provided no account of which sorts of states have actual intrinsic value and which have merely virtual intrinsic value.[33] I plead guilty, for I have not attempted to supply even part of a substantive axiology. For purposes of illustration, I have suggested that pleasure as such may have merely virtual intrinsic value whereas undeserved pleasure in the bad is perhaps actually intrinsically bad. But I certainly have not argued for this view; that would be an entirely separate task. My failure to undertake this task, however, affords no reason to deny the legitimacy of the distinction between actual and virtual intrinsic value. Given this distinction, one can envision a gamut of axiologies, ranging from the extremely atomistic (according to which all states have actual intrinsic value, none have merely virtual intrinsic value) to the extremely holistic (according to which the only state with actual intrinsic value is the actual world). I suspect that the true axiology lies somewhere in between (so that while some states have actual intrinsic value – and the intrinsic value of conjunctions of these states is to be computed simply by summation – other states have merely virtual intrinsic value), but this claim cannot be investigated here.

A third objection is that the present approach violates what is sometimes called the principle of universality.[34] This principle may be put as follows: the intrinsic value of a part of an intrinsically valuable whole does not depend on the other parts of the whole. It is to be contrasted with the principle of conditionality, which may be put as follows:

---

[33] Chisholm mentions this in Chisholm (1986), pp. 81-82, n. 22 [* p. 313 of this volume].
[34] Cf. Lemos (1994), pp. 32-33.

the intrinsic value of a part of an intrinsically valuable whole does depend on the other parts of the whole. Lemos is one author who declares the former principle true and the latter false, and he accuses Ross of being committed to the latter rather than the former. But whether or not this is true of Ross, it is not true of me. I agree that the intrinsic value – that is, the actual intrinsic value – of a state cannot be conditional on the state's circumstances. My claim is that states that have merely virtual intrinsic value have no actual intrinsic value, and that this is so regardless of the circumstances of such states.

A final and, to my mind, much stronger and more interesting objection is one that has been offered by Lemos:

> The chief difficulty with Ross's view [that pleasure as such is merely prima facie intrinsically good] is that it misses what apparently makes so offensive the prosperity of the wicked... What is it that makes the wicked man's being happy...so offensive? I suggest that it is offensive precisely because we think that the wicked man has a *good* that he deserves not to have. Contrary to Ross, the judgment that the prosperity of the wicked is not good...presupposes the judgment that his being pleased *is* good; it is a good that is contrary to what he deserves.[35]

There is undoubtedly something correct in this. When a wicked man prospers, he benefits undeservedly, and *that* is what seems so offensive. And it is not just he who sees his prosperity as being beneficial to him; we do too. But I think we should resist saying that there is anything *intrinsically good* involved in this benefit. His prosperity, being beneficial to him, might be said to be *good for him*; but it doesn't follow that it is *good simpliciter*. It is not clear to me just how this distinction is to be understood, but we can gain at least partial insight into it by recalling that, when it is said that contemplation of what is good requires that one favor it, it is *ethical* requirement that is at issue. Consider Wilfred's being both wicked and happy. Lemos would say that contemplation of this state (and this state alone) requires that we disfavor it, whereas contemplation of Wilfred's being happy requires that we favor it. Although I am prepared (for the sake of argument) to accept the former claim, I reject the latter. Notice that it is surely rational (in some sense) for Wilfred himself not just to prefer his being happy to his not being happy, but also to prefer his being wicked and happy to his being wicked and not happy; but even he (like everyone else) is ethically required to prefer his being wicked and not happy to his being wicked and happy. This indicates that the sort of rationality at issue here is distinct from ethical requirement, and I can therefore agree that it is rational for the wicked Wilfred to prefer his being happy to his not being happy (and that it is at least partly in virtue of this fact that his being happy is good *for him*), without agreeing that everyone is ethically required to prefer his being happy to his not being happy (and so without agreeing that his being happy is intrinsically good).

---

[35] Lemos (1994), pp. 43-44. See also Goldstein (1989), p. 269, for a very similar point.

7. EXTENSIONS OF THE ACCOUNT

In a previous paper I made no mention of "virtual" intrinsic value, but I did talk of the possibility of certain states being "evaluatively incomplete," that is, entirely lacking in (actual) intrinsic value.[36] I applied this to cases where organic unities have been diagnosed, but I applied it to other cases too. Suppose that a very simple form of hedonism were true, according to which (roughly) all pleasures and only pleasures are intrinsically good, and the goodness of pleasures cannot be defeated by their being embedded in any wider context. Even on such a supposition, it seems to me, one should hesitate to say that a state consisting in someone's being pleased is itself actually intrinsically good. After all, what is good is presumably good to a certain determinate extent (I suppose that this could be challenged, perhaps on the basis that some goods are incommensurable with others, but I will let this pass), and to *what* extent is the state in question good? Well, it depends, doesn't it, on the context or situation at hand. If Peter is pleased to degree 10 (to use numbers again in a purely *ad hoc* fashion) then, on this simple hedonism, there would be reason to say that the state consisting in his being pleased to degree 10 is intrinsically good and, more particularly, good to degree 10. But what of the state consisting simply in Peter's being pleased? *How* good is *it*? My answer, of course, is that this state is, even on the simple hedonism being presupposed, not *actually* good at all; rather, it is merely *virtually* good (to degree 10).[37]

We may treat disjunctive states (such as Peter's being either pleased or displeased) similarly. Committed to the view that all states have actual intrinsic value, Chisholm has gone to great lengths to establish the proper criterion for determining the intrinsic value of disjunctive states.[38] His view, whose details I shall not provide here, has the strange consequence that a disjunctive state consisting in the disjunction of a very good state and a slightly bad state has the same intrinsic value as the disjunction of a slightly good state and a very bad state (namely, nil); it also implies that the disjunction of a slightly good state and another slightly good state has the same intrinsic value as the disjunction of a slightly good state and a very good state (namely, the value of the slightly good state). All of this can be avoided by noting that a disjunctive state cannot exist without one or other of its disjuncts existing and that, therefore, we can ascribe one or more *virtual* intrinsic values to the disjunctive state (these values corresponding to the actual intrinsic values, if any, of its existing disjuncts) without having to try to settle on a single actual intrinsic value for the disjunction itself.

---

[36] Zimmerman (1983). In that paper I took abstract states of affairs to be the bearers of intrinsic value, contrary to the position I take in this paper.

[37] That is, it has a virtual intrinsic value of 10. Note that this is consistent with saying that it may *also* have a *different* virtual intrinsic value (since it may be part of some state, *other* than that of Peter's being pleased to degree 10, that has an actual intrinsic value other than 10).

[38] Chisholm (1975) [* pp. 229-39 of this volume].

8. BASIC INTRINSIC VALUE

When responding to the first objection in Section 6, I proposed that the principle of organic unities be reformulated as follows:

> The actual intrinsic value of a state must not be assumed to be the same as the sum of the actual intrinsic values of those states that are a part of it and have such value.

But in fact this won't quite do, either. When trying to determine the actual intrinsic value of a suitably "wide" state, one cannot do this simply by summing up the actual intrinsic values, if any, of its "component" states, for this would result in serious distortion. Suppose that S1 is intrinsically good to degree 10, S2 to degree 20, and S3 to degree 30. Then, it seems, we may talk not only of S1 and S2 and S3 each individually being a component of S1&S2&S3, but also of S1&S2 and S1&S3 and S2&S3 likewise being components of S1&S2&S3. Perhaps we can also talk of S1&S2&S3 being a component of itself. The sum value of all these components is 240 – far higher than the value of 60 that we are intuitively moved to assign to S1&S2&S3. Our intuition is based, of course, on the tacit assumption that S1 and S2 and S3 are not just evaluatively complete but *basically* so, whereas this is not true of the other parts of S1&S2&S3.

I think that the notion of basic intrinsic value can be fairly straightforwardly understood, once it is acknowledged that not all states have actual intrinsic value. We can say the following:

> S has basic intrinsic value =df. S has actual intrinsic value, and no proper part of S has actual intrinsic value.[39]

On this basis, we can reformulate the relevant principle of organic unities yet again, as follows:

> The actual intrinsic value of a state must not be assumed to be the same as the sum of the basic actual intrinsic values of those states that are a part of it and have such value.[40]

I cannot claim to have shown this principle false, but I hope that I have given good reason to doubt that it is true.[41]

---

[39] This isn't quite adequate, since every state seems infinitely temporally divisible. But this is a complication that I shall ignore here.

[40] On basic intrinsic value, and how this notion can be put to a variety of uses, see Feldman (2000) [* pp. 379-400 of this volume].

[41] My thanks to Krister Bykvist, Irwin Goldstein, Joshua Hoffman, Noah Lemos, Howard Sobel, and some anonymous referees for very penetrating and helpful comments on earlier drafts.

BIBLIOGRAPHY

Blanshard, Brand (1961). *Reason and Goodness*. London: George Allen and Unwin.
Brentano, Franz (1969). *The Origin of Our Knowledge of Right and Wrong*, translated by Roderick M. Chisholm and Elizabeth H. Schneewind. New York: Humanities Press.
Butchvarov, Panayot (1989). *Skepticism in Ethics*. Indianapolis: Indiana University Press.
Chisholm, Roderick M. (1964). "The Ethics of Requirement." *American Philosophical Quarterly* 1: 147-53.
_____ (1968-69). "The Defeat of Good and Evil." *Proceedings of the American Philosophical Association* 42: 21-38.
_____ (1975). "The Intrinsic Value in Disjunctive States of Affairs." *Noûs* 9: 295-308 [* pp. 229-39 of this volume].
_____ (1986). *Brentano and Intrinsic Value*. Cambridge: Cambridge University Press.
Dancy, Jonathan (1993). *Moral Reasons*. Oxford: Blackwell.
Ewing, A. C. (1948). *The Definition of Good*. London: Routledge and Kegan Paul.
Feldman, Fred (1986). *Doing the Best We Can*. Dordrecht: D. Reidel.
_____ (2000). "Basic Intrinsic Value." *Philosophical Studies* 99: 319-46 [* pp. 379-400 of this volume].
Goldstein, Irwin (1989). "Pleasure and Pain: Unconditional, Intrinsic Values." *Philosophy and Phenomenological Research* 50: 255-76.
_____ (1996). "Pain's Intrinsic Badness and Ethical Anti-Realism." In *The Problematic Reality of Values*, edited by Jan Bransen and Marc Slors (Assen: Van Gorcum): 43-49.
Lemos, Noah (1994). *Intrinsic Value*. Cambridge: Cambridge University Press.
Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
_____ (1912). *Ethics*. Oxford: Oxford University Press.
Paton, H. J. (1942). "The Alleged Independence of Goodness." In *The Philosophy of G. E. Moore*, edited by P. A. Schilpp (Chicago: Northwestern University): 113-34.
Ross, W. D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
_____ (1939). *Foundations of Ethics*. Oxford: Clarendon Press.
Zimmerman, Michael J. (1983). "Evaluatively Incomplete States of Affairs." *Philosophical Studies* 43: 211-24.

BIBLIOGRAPHY

Allen, Julie (2003). "G. E. Moore and the Principle of Organic Unity." *Journal of Value Inquiry*, 37: 329-39.

Alm, David (2004). "Atomism about Value." *Australasian Journal of Philosophy*, 82: 312-31.

Anderson, C. Anthony (1997). "Chisholm and the Logic of Intrinsic Value." In *The Philosophy of Roderick M. Chisholm*, edited by L. E. Hahn (Chicago: Open Court): 407-29.

Anderson, Elizabeth (1993). *Value in Ethics and Economics*. Cambridge, Mass.: Harvard University Press.

Åqvist, Lennart (1968). "Chisholm-Sosa Logics of Intrinsic Betterness and Value." *Noûs*, 2: 253-70.

Arrhenius, Gustaf (2000). "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy*, 16: 247-66.

Attfield, Robin (1987). *A Theory of Value and Obligation*. New York: Croom Helm.

_____ (1991). *The Ethics of Environmental Concern*. Athens: University of Georgia Press.

Audi, Robert (1997a). "Intrinsic Value and Moral Obligation." *Southern Journal of Philosophy*, 35: 135-54.

_____ (1997b). *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press.

_____ (1998). "The Axiology of Moral Experience." *Journal of Ethics*, 2: 355-75.

_____ (2003a). "Intrinsic Value and Reasons for Action." *Southern Journal of Philosophy*, suppl. vol. 41: 30-56.

_____ (2003b). "Intrinsic Value, Inherent Value, and Experience: A Reply to Stephen Barker." *Southern Journal of Philosophy*, 61: 323-27.

_____ (2004). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.

Baldwin, Thomas (2003). "The Indefinability of Good." *Journal of Value Inquiry*, 37: 313-28.

Barker, Stephen (2003). "The Experiential Thesis: Audi on Intrinsic Value." *Southern Journal of Philosophy*, suppl. vol. 41: 57-61.

Baron, Marcia W. (1986). "On Admirable Immorality." *Ethics*, 96: 557-66.

Baron, Marcia W., et al. (1997). *Three Methods of Ethics*. Oxford: Blackwell.

Baylis, Charles A., (1958). "Grading, Values, and Choice." *Mind*, 67: 485-501.

Beardsley, Elizabeth L. (1957). "Moral Worth and Moral Credit." *Philosophical Review*, 66: 304-28.

Berlin, Isaiah (1969). *Four Essays on Liberty*. Oxford: Oxford University Press.

Bernstein, M. (2001). "Intrinsic Value." *Philosophical Studies*, 102: 329-43.

Blackburn, Simon (1984). *Spreading the Word*. Oxford: Oxford University Press.

Blanshard, Brand (1961). *Reason and Goodness*. London: Allen and Unwin.
Bloomfield, Paul (2003). "Partially Re-Humanized Ethics: Comments on Butchvarov." *Southern Journal of Philosophy*, suppl. vol. 41: 184-89.
Bradley, Ben (1998). "Extrinsic Value." *Philosophical Studies*, 91: 109-26.
_____ (2001). "The Value of Endangered Species." *Journal of Value Inquiry*, 35: 43-58.
_____ (2002). "Is Intrinsic Value Conditional?" *Philosophical Studies*, 107: 23-44.
Brandt, Richard B. (1959). *Ethical Theory*. Englewood Cliffs: Prentice-Hall.
_____ (1963). *Value and Obligation.* New York: Harcourt, Brace, and World.
_____ (1979). *A Theory of the Good and the Right*. Oxford: Clarendon Press.
Brännmark, Johan (2001). "Good Lives: Parts and Wholes." *American Philosophical Quarterly*, 38: 221-31.
Broome, John (1991). *Weighing Goods*. Oxford: Blackwell.
_____ (1999a). *Ethics out of Economics*. Cambridge: Cambridge University Press.
_____ (1999b). "Normative Requirements." *Ratio*, 12: 398-419.
_____ (2004). *Weighing Lives*. Oxford: Oxford University Press.
Brülde, Bengt (1998). *The Human Good.* Göteborg: Acta Universitatis Gothoburgensis.
Butchvarov, Panayot (1989). *Skepticism in Ethics*. Indianapolis: Indiana University Press.
_____ (2003). "Ethics Dehumanized." *Southern Journal of Philosophy*, suppl. vol. 41: 165-83.
Bykvist, Krister (1998). *Changing Preferences: A Study in Preferentialism.* Uppsala: Acta Universitatis Uppsaliensis.
Carlson, Erik (1995). *Consequentialism Reconsidered.* Dordrecht: Kluwer.
_____ (1997). "A Note on Moore's Organic Unities." *Journal of Value Inquiry*, 31: 55-59.
_____ (2000). "Organic Unities, Weak Lexicality, and the Additivity of Intrinsic Value." In *Value and Choice*, edited by Wlodek Rabinowicz (Lund: Lund Philosophy Reports): 65-89.
_____ (2003). "On Some Recent Examples of Intransitive Betterness." In *Logic, Law, Morality*, edited by Krister Segerberg and Rysiek Sliwinski (Uppsala: Department of Philosophy): 181-95.
Carson, Thomas L. (2000). *Value and the Good Life*. South Bend: University of Notre Dame Press.
Carter, Robert Edgar (1968). "The Importance of Intrinsic Value." *Philosophy and Phenomenological Research*, 28: 567-77.
_____ (1974). "Intrinsic Value and the Intrinsic Valuer." *Philosophy and Phenomenological Research*, 34: 504-14.
Castañeda, Hector-Neri (1969). "Ought, Value, and Utilitarianism." *American Philosophical Quarterly*, 6: 257-75.
Chang, Ruth, (ed.) (1997a). *Incommensurability, Incomparability, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
_____ (2002). "The Possibility of Parity." *Ethics*, 112: 659-88.

Cheney, Jim (1992). “Intrinsic Value in Environmental Ethics: Beyond Subjectivism and Objectivism.” *Monist*, 75: 227-35.
Chisholm, Roderick M. (1968-69). “The Defeat of Good and Evil.” *Proceedings and Addresses of the American Philosophical Association*, 42: 21-38.
_____ (1974). “Practical Reason and the Logic of Requirement.” In *Practical Reason*, edited by Stephan Körner (Oxford: Blackwell): 1-17.
_____ (1986). *Brentano and Intrinsic Value*. Cambridge: Cambridge University Press.
Chisholm, Roderick M. and Sosa, Ernest (1966a). “Intrinsic Preferability and the Problem of Supererogation.” *Synthese*, 16: 321-31.
_____ (1966b). “On the Logic of ‘Intrinsically Better’.” *American Philosophical Quarterly*, 3: 244-49.
Conee, Earl (1982). “Instrumental Value without Intrinsic Value?” *Philosophia*, 11: 345-59.
Dancy, Jonathan (1993). *Moral Reasons.* Oxford: Blackwell.
_____ (1999). “On the Logical and Moral Adequacy of Particularism.” *Theoria*, 65: 212-24.
_____ (2003). “Are There Organic Unities?” *Ethics*, 113: 629-50.
_____ (2004). *Ethics without Principles*. Oxford: Oxford University Press.
Danielsson, Sven (1997). “Harman’s Equation and the Additivity of Intrinsic Value.” *Uppsala Philosophical Studies*, 46: 23-34.
_____ (2001). “The Supervenience of Intrinsic Value.” *Uppsala Philosophical Studies*, 50: 93-103.
D’Arms, Justin and Jacobson, Daniel (2000). “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions.” *Philosophy and Phenomenological Research*, 61: 65-90.
Darwall, Stephen (1977). “Two Kinds of Respect.” *Ethics*, 88: 36-49.
_____ (2003). “Moore, Normativity, and Intrinsic Value.” *Ethics*, 113: 468-89.
DePaul, Michael R. (2002). “A Half Dozen Puzzles regarding Intrinsic Attitudinal Hedonism.” *Philosophy and Phenomenological Research*, 65: 629-35.
Dreier, Jamie (2003). “Gibbard and Moore.” *Southern Journal of Philosophy*, suppl. vol. 41: 158-64.
Ducasse, C. J. (1968). “Intrinsic Value.” *Philosophy and Phenomenological Research*, 28: 410-12.
Duncan-Jones, Austin (1958). “Intrinsic Value: Some Comments on the Work of G. E. Moore,” *Philosophy*, 33: 240-73.
_____ (1966). “Good Things and Good Thieves.” *Analysis*, 26: 113-18.
Edwards, Rem B. (1979). *Pleasures and Pains*. Ithaca: Cornell University Press.
Egonsson, Dan (1998). *Dimensions of Dignity*. Dordrecht: Kluwer.
Elliot, Robert (1992). “Intrinsic Value, Environmental Obligation and Naturalness.” *Monist*, 75: 138-60.
Ewing, A. C. (1973). *Value and Reality*. London: George Allen and Unwin.
Feldman, Fred (1986). *Doing the Best We Can*. Dordrecht: Reidel.
_____ (1992). *Confrontations with the Reaper*. Oxford: Oxford University Press.

_____ (1997). *Utilitarianism, Hedonism, and Desert*. Cambridge: Cambridge University Press.

_____ (2002a). "The Good Life: A Defense of Attitudinal Hedonism." *Philosophy and Phenomenological Research*, 65: 604-28.

_____ (2002b). "Comments on Two of DePaul's Puzzles." *Philosophy and Phenomenological Research*, 65: 636-39.

Findlay, J. N. (1963a). *Language, Mind, and Value*. London: Allen and Unwin.

_____ (1963b). *Meinong's Theory of Objects and Values*. Oxford: Oxford University Press.

_____ (1968). *Values and Intentions*. London: Allen and Unwin.

Finnis, John (1983). *Fundamentals of Ethics*. Oxford: Clarendon Press.

Flanagan, Owen (1986). "Admirable Immorality and Admirable Imperfection." *Journal of Philosophy*, 83: 41-60.

Foot, Philippa (1961). "Goodness and Choice." *Proceedings of the Aristotelian Society*, suppl. vol. 35: 45-60.

_____ (1978). *Virtues and Vices*. Berkeley: University of California Press.

_____ (1985). "Utilitarianism and the Virtues." *Mind*, 94: 196-209.

Frankena, William K. (1973). *Ethics*, second edition. Englewood Cliffs: Prentice Hall.

Frankfurt, Harry G. (1999). *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

Freeman, James B. (1973). "Fairness and the Value of Disjunctive Actions." *Philosophical Studies*, 24: 105-11.

Fried, Charles (1970). *An Anatomy of Values*. Cambridge, Mass.: Harvard University Press.

Frondizi, Robert (1963). *What Is Value?* La Salle: Open Court.

Garcia, J. L. A. (1987). "Goods and Evils." *Philosophy and Phenomenological Research*, 47: 385-412.

Gardiner, P. L. (1964). "Pain and Evil." *Proceedings of the Aristotelian Society*, suppl. vol. 38: 107-24.

Gaus, Gerald F. (1990). *Value and Justification*. Cambridge: Cambridge University Press.

Geach, Peter (1956). "Good and Evil." *Analysis*, 17: 33-42.

_____ (1977). *The Virtues*. New York: Cambridge University Press.

Gibbard, Alan (1990). *Wise Choices, Apt Feelings*. Cambridge, Mass: Harvard University Press.

_____ (2003). "Normative Properties." *Southern Journal of Philosophy*, suppl. vol. 41: 141-57.

Goldstein, Irwin (1989). "Pleasure and Pain: Unconditional, Intrinsic Values." *Philosophy and Phenomenological Research*, 50: 255-76.

_____ (1996). "Pain's Intrinsic Badness and Ethical Anti-Realism." In *The Problematic Reality of Values*, edited by Jan Bransen and Marc Slors (Assen: Van Gorcum): 43-49.

_____ (2002). "The Good's Magnetism and Ethical Realism." *Philosophical Studies*, 108: 1-14.

_____ (2003). "Malicious Pleasure Evaluated: Is Pleasure an Unconditional Good?" *Pacific Philosophical Quarterly*, 84: 24-31.

Grice, Paul (1991). *The Conception of Value*. Oxford: Clarendon Press.

Griffin, James (1986). *Well-Being*. Oxford: Clarendon Press.

Hall, Everett (1961). *Our Knowledge of Fact and Value.* Chapel Hill: University of North Carolina Press.

Hansson, Bengt (1968a). "Fundamental Axioms for Preference Relations." *Synthese*, 18: 423-42.

_____ (1968b). "Choice Structures and Preference Relations." *Synthese*, 18: 443-58.

Hansson, Sven Ove (2001**).** *The Structure of Values and Norms.* Cambridge: Cambridge University Press.

Hare, Richard M. (1964). "Pain and Evil." *Proceedings of the Aristotelian Society*, suppl. vol. 38: 91-106.

Hargrove, Eugene C. (1992). "Weak Anthropocentric Intrinsic Value." *Monist*, 75: 183-207.

Harman, Gilbert (2000). *Explaining Values*. Oxford: Oxford University Press.

Harman, Gilbert and Thomson, Judith Jarvis (1996). *Moral Realism and Moral Objectivity*. Oxford: Blackwell.

Harris, George W. (2001). "Value Vagueness, Zones of Incomparability, and Tragedy." *American Philosophical Quarterly*, 38: 155-76.

Haybron, Daniel M. (1999). "Evil Characters." *American Philosophical Quarterly*, 36: 137-48.

Holtug, Nils (2003). "Good for Whom?" *Theoria*, 69: 3-20.

Hooker, Brad and Little, Margaret (ed.) (2000). *Moral Particularism*. Oxford: Clarendon Press.

Hospers, John (1961). *Human Conduct.* New York: Harcourt, Brace, and World.

Hurka, Thomas (1998). "Two Kinds of Organic Unity." *Journal of Ethics*, 2: 299-320.

_____ (2000). *Virtue, Vice, and Value*. New York: Oxford University Press.

_____ (2002). "The Common Structure of Virtue and Desert." *Ethics*, 112: 6-31.

_____ (2003). "Moore in the Middle." *Ethics*, 113: 599-628.

Jackson, Frank (2003). "Cognitivism, A Priori Deduction, and Moore." *Ethics*, 113: 557-75.

Johnston, Mark (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society*, suppl. vol. 63: 139-74.

Kagan, Shelly (1988). "The Additive Fallacy." *Ethics*, 99: 5-31.

_____ (1998). "Rethinking Intrinsic Value." *Journal of Ethics*, 2: 277-97.

Keller, Simon (2000). "How Do I Love Thee? Let Me Count the Properties." *American Philosophical Quarterly*, 37: 163-73.

Korsgaard, Christine M. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

_____ (1997). "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Kraut (Oxford: Oxford University Press): 215-54.

_____ (1998). "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer and Schneewind." *Ethics*, 109: 49-66.

Kupperman, Joel J. (1999). *Value...And What Follows*. Oxford: Oxford University Press.

Lemos, Noah (1994). *Intrinsic Value*. Cambridge: Cambridge University Press.

_____ (1998). "Organic Unities." *Journal of Ethics*, 2: 321-37.

Lemos, Ramon (1991). "Bearers of Value." *Philosophy and Phenomenological Research*, 51: 873-89.

Lewis, David (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society*, suppl. vol. 63: 113-37.

Lippert-Rasmussen, Kasper (2003). "Measuring the Disvalue of Inequality over Time." *Theoria*, 69: 32-45.

Magendanz, Douglas (2003). "Conflict and Complexity in Value Theory." *Journal of Value Inquiry*, 37: 443-53.

Magnell, Thomas (1993a). "Evaluations as Assessments, Part I: Properties and Their Signifiers." *Journal of Value Inquiry*, 27: 1-11.

_____ (1993b). "Evaluations as Assessments, Part II: Distinguishing Assertions and Instancing Good of a Kind." *Journal of Value Inquiry*, 27: 151-63.

Mayerfeld, Jamie (1999). *Suffering and Moral Responsibility*. Oxford: Oxford University Press.

McCarty, Richard (1994). "Are There 'Contra-Moral Virtues'?" *Metaphilosophy*, 25: 362-75.

Montefiore, Alan (1961). "Goodness and Choice." *Proceedings of the Aristotelian Society*, suppl. vol. 35: 61-80.

Mulligan, Kevin (1998). "From Appropriate Emotions to Values." *Monist*, 81: 161-88.

Murdoch, Iris (1991). *The Sovereignty of Good*. London: Routledge.

Nagel, Thomas (1979). *Mortal Questions*. Cambridge: Cambridge University Press.

_____ (1989). *The View from Nowhere*. New York: Oxford University Press.

Norton, Bryan G. (ed.) (1986). *The Preservation of Species*. Princeton: Princeton University Press.

_____ (1992). "Epistemology and Environmental Values." *Monist*, 75: 208-26.

Nozick, Robert (1981). *Philosophical Explanations*. Oxford: Oxford University Press.

_____ (1993). *The Nature of Rationality.* Princeton: Princeton University Press.

Nussbaum, Martha C. (1984). *The Fragility of Goodness*. Cambridge: Cambridge University Press.

Oddie, Graham (2001a). "Axiological Atomism." *Australasian Journal of Philosophy*, 79: 313-32.

_____ (2001b). "Recombinant Values." *Philosophical Studies*, 106: 259-92.

Oldfield, Edward (1977). "An Approach to a Theory of Intrinsic Value." *Philosophical Studies*, 32: 233-49.

Olson, Jonas (2003). "Revisiting the Tropic of Value: Reply to Rabinowicz and Rønnow-Rasmussen." *Philosophy and Phenomenological Research*, 67: 412-22.

_____ (2004a). "Buck-Passing and the Wrong Kind of Reasons." *Philosophical Quarterly*, 54: 295-300.

_____ (2004b). "Intrinsicalism and Conditionalism about Final Value." *Ethical Theory and Moral Practice*, 7: 31-52.

O'Neill, John (1992). "The Varieties of Intrinsic Value." *Monist*, 75: 119-37.

Österberg, Jan (1996). "Value and Existence." *Uppsala Philosophical Studies*, 45: 94-107.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.

_____ (1986). "Overpopulation and the Quality of Life." In *Applied Ethics*, edited by Peter Singer (Oxford: Oxford University Press, 1986): 145-64.

Parsons, Charles (1970). "Axiomatization of Åqvist's CS Logics." *Theoria*, 36: 43-64.

Pasquerella, Lynn (1985). "Brentano and Organic Unities." In *From Bolzano to Wittgenstein*, edited by J. C. Nyiri (Vienna: Hölder-Pichler-Tempsky): 121-31.

Pastin, Mark (1975). "The Reconstruction of Value." *Canadian Journal of Philosophy*, 5: 375-93.

Patton, Thomas E. and Ziff, Paul (1964). "On Vendler's Grammar of 'Good'," *Philosophical Review*, 73: 528-37.

Paul, E. F., et al. (ed.) (1992). *The Good Life and the Human Good*. Cambridge: Cambridge University Press.

Perrett, Roy W. (1997). "Preferring More Pain to Less." *Philosophical Studies*, 93: 213-26.

_____ (2003). "Recognizing and Reacting to Value." *Journal of Value Inquiry*, 37: 51-58.

Persson, Ingmar (1996). "Benevolence, Identification and Value." *Uppsala Philosophical Studies*, 45: 58-71.

_____ (1997). "Ambiguities in Feldman's Desert-adjusted Values." *Utilitas*, 9: 319-27.

_____ (2003). "The Badness of Unjust Inequality." *Theoria*, 69: 109-24.

Pettit, Philip (1993). "Consequentialism." In *A Companion to Ethics*, edited by Peter Singer (Oxford: Blackwell): 230-40.

Phillips, David (2003). "Thomson and the Semantic Argument against Consequentialism." *Journal of Philosophy*, 100: 475-86.

Pitcher, George (1970). "The Awfulness of Pain." *Journal of Philosophy*, 67: 481-92.

Quinn, Philip C. (1978). *Divine Commands and Moral Requirements*. Oxford: Clarendon Press.

Quinn, Warren S. (1974). "Theories of Intrinsic Value." *American Philosophical Quarterly* 11: 123-32.

_____ (1990). "The Puzzle of the Self-Torturer." *Philosophical Studies*, 59: 79-90.

Rabinowicz, Wlodek (ed.) (2000). *Value and Choice: Some Common Themes in Decision Theory and Moral Philosophy*, vol. 1. Lund: Lund Philosophy Reports.
_____ (ed.) (2001a) *Value and Choice: Some Common Themes in Decision Theory and Moral Philosophy*, vol. 2. Lund: Lund Philosophy Reports.
_____ (2001b) "Prioritarianism and Uncertainty: On the Interpersonal Addition Theorem and the Priority View." In *Exploring Practical Philosophy: From Action to Values*, edited by D. Egonsson, J. Josefsson, B. Petersson, and T. Rønnow-Rasmussen (Aldershot: Ashgate): 139-65.
_____ (2002). "Prioritarianism for Prospects." *Utilitas*, 14: 2-21.
_____ (2003a). "Ryberg's Doubts about Higher Pleasures – Put to Rest?" *Ethical Theory and Moral Practice*, 6: 231-35.
_____ (2003b). "The Size of Inequality and Its Badness: Some Reflections around Temkin's *Inequality*." *Theoria*, 69: 60-84.
Rabinowicz, Wlodek and Österberg, Jan (1996). "Value Based on Preferences." *Economics and Philosophy*, 12: 1-27.
Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (ed.) (2003). *Patterns of Value*, vol 1. Lund: Lund Philosophy Reports.
_____ (ed.) (2004a). *Patterns of Value*, vol 2. Lund: Lund Philosophy Reports.
_____ (2004b). "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics*, 114: 391-423.
Rachels, Stuart (2001). "A Set of Solutions to Parfit's Problems." *Noûs*, 35: 214-38.
_____ (2003). "A Defense of Two Optimistic Claims in Ethical Theory." *Philosophical Studies*, 112: 1-30.
Raz, Joseph (1999). *Engaging Reason*. Oxford: Oxford University Press.
_____ (2001). *Value, Respect and Attachment.* Cambridge: Cambridge University Press.
Regan, Donald H. (2002). "The Value of Rational Nature." *Ethics*, 112: 267-91.
_____ (2003). "How to Be a Moorean." *Ethics*, 113: 651-77.
Regan, Tom, (ed.) (1984). *Earthbound: New Introductory Essays in Environmental Ethics*. Philadelphia: Temple University Press.
_____ (1986). *Bloomsbury's Prophet*. Philadelphia: Temple University Press.
_____ (1992). "Does Environmental Ethics Rest on a Mistake?" *Monist*, 75: 161-82.
Rescher, Nicholas (1967). "Semantic Foundations for the Logic of Preference." In *The Logic of Decision and Action*, edited by Nicholas Rescher (Pittsburgh: University of Pittsburgh Press): 37-62.
_____ (1993). *The Validity of Values*. Princeton: Princeton University Press.
Rolston, Holmes, III (1992). "Disvalues in Nature." *Monist*, 75: 250-78.
Rønnow-Rasmussen, Toni (2002a). "Hedonism, Preferentialism, and Value Bearers." *Journal of Value Inquiry*, 36: 463-72.
_____ (2002b). "Instrumental Values – Strong and Weak." *Ethical Theory and Moral Practice*, 5: 23-43.
Rosati, Connie (2003). "Agency and the Open Question Argument." *Ethics*, 113: 490-527.

Ryberg, Jesper (2002). "Higher and Lower Pleasures – Doubts on Justification." *Ethical Theory and Moral Practice*, 5: 415-29.
Scanlon, T. M. (1998). *What We Owe to Each Other.* Cambridge, Mass.: Harvard University Press.
Schaber, Peter (1999). "Value Pluralism: Some Problems." *Journal of Value Inquiry*, 33: 71-78.
Schroth, Jörg (2003). "Particularism and Universalizability." *Journal of Value Inquiry*, 37: 455-61.
Schumm, George (1987). "Transitivity, Preference and Indifference." *Philosophical Studies*, 52: 435-37.
Sinnott-Armstrong, Walter (2003). "For Goodness' Sake." *Southern Journal of Philosophy*, suppl. vol. 41: 83-91.
Slote, Michael (1983). *Goods and Virtues*. Oxford: Clarendon Press.
Smith, Holly M. (1991). "Varieties of Moral Credit and Moral Worth." *Ethics*, 101: 279-303.
Smith, Michael (2003). "Neutral and Relative Value after Moore." *Ethics*, 113: 576-98.
Smith, Tara (1998). "Intrinsic Value: Look-Say Ethics." *Journal of Value Inquiry*, 32: 539-53.
Sobel, David (2002). "Varieties of Hedonism." *Journal of Social Philosophy*, 33: 240-56.
Sobel, J. Howard (1970). "Utilitarianisms: Simple and General." *Inquiry*, 13: 394-449.
Stocker, Michael (1990). *Plural and Conflicting Values*. Oxford: Clarendon Press.
Stratton-Lake, Philip (2000). *Kant, Duty, and Moral Worth.* London: Routledge.
Sturgeon, Nicholas L. (1996). "Anderson on Reason and Value." *Ethics*, 106: 509-24.
_____ (2003). "Moore and Ethical Naturalism." *Ethics*, 113: 528-56.
Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
Swanton, Christine (1995). "Profiles of the Virtues." *Pacific Philosophical Quarterly*, 27: 335-44.
Tännsjö, Torbjörn (1996). "Classical Hedonistic Utilitarianism." *Philosophical Studies*, 81: 97-115.
_____ (1998). *Hedonistic Utilitarianism.* Edinburgh: Edinburgh University Press.
Taylor, Paul (1961). *Normative Discourse*. New York: Prentice-Hall.
Temkin, Larry S. (1987). "Intransitivity and the Mere Addition Paradox." *Philosophy and Public Affairs*, 16: 138-87.
_____ (1993). *Inequality.* Oxford: Oxford University Press.
_____ (1994). "Weighted Goods: Some Questions and Comments." *Philosophy and Public Affairs*, 23: 361-63.
_____ (1996). "A Continuum Argument for Intransitivity." *Philosophy and Public Affairs*, 25: 175-210.

_____ (1997). "Rethinking the Good, Moral Ideals and the Nature of Practical Reasoning." In *Reading Parfit*, edited by Jonathan Dancy (Oxford: Blackwell): 290-344.

_____ (2003a). "Personal versus Impersonal Principles: Reconsidering the Slogan." *Theoria*, 69: 21-31.

_____ (2003b). "Determining the Scope of Egalitarian Concern: A Partial Defense of Complete Lives Egalitarianism." *Theoria*, 69: 46-59.

_____ (2003c). "Measuring Inequality's Badness: Does Size Matter? If So, How, If Not, What Does?" *Theoria*, 69: 85-108.

_____ (2003d). "Exploring the Roots of Egalitarian Concerns." *Theoria*, 69: 125-51.

Thomson, Judith Jarvis (1994). "Goodness and Utilitarianism." *Proceedings and Addresses of the American Philosophical Association*, 67: 7-21.

_____ (2001). *Goodness and Advice*. Princeton: Princeton University Press.

_____ (2003a). "The Legacy of *Principia*." *Southern Journal of Philosophy*, suppl. vol. 41: 62-82.

_____ (2003b). "Reply to Sinnott-Armstrong." *Southern Journal of Philosophy*, suppl. vol. 41: 92-94.

Tolhurst, William (1983). "On the Nature of Intrinsic Value." *Philosophical Studies*, 43: 383-95.

Vallentyne, Peter and Kagan, Shelly (1997). "Infinite Value and Finitely Additive Value Theory." *Journal of Philosophy*, 94: 5-26.

Velleman, J. David (1999). "Love as a Moral Emotion." *Ethics*, 109: 338-75.

Vendler, Zeno (1963). "The Grammar of Goodness." *Philosophical Review*, 72: 446-65.

Wallace, James D. (1978). *Virtues and Vices*. Ithaca: Cornell University Press.

Weston, Anthony (1985). "Beyond Intrinsic Value: Pragmatism in Environmental Ethics." *Environmental Ethics*, 7: 321-39.

_____ (1995). "Between Means and Ends." *Monist*, 75: 236-49.

Wielenberg, E. (1998). "Goodness without Qualification." *Journal of Value Inquiry*, 32: 93-104.

Williams, Bernard (1972). *Morality*. New York: Harper and Row.

Willigenburg, T. van (1999). "Is the Consumer Always Right? Intrinsic Value and Subject-Relativity." In *Ethics and the Market*, edited by R. Norman (Aldershot: Ashgate Press): 26-46.

Wolf, Susan (1982). "Moral Saints." *Journal of Philosophy*, 79: 419-39.

_____ (1997). "Happiness and Meaning: Two Aspects of the Good Life." *Social Philosophy and Policy*, 14: 207-25.

Wright, Georg Henrik von (1963). *The Varieties of Goodness*. London: Routledge and Kegan Paul.

_____ (1972). "The Logic of Preference Reconsidered." *Theory and Decision*, 3: 140-69.

Yarnell, Patrick H. (2001). "The Intrinsic Goodness of Pain, Anguish, and the Loss of Pleasure." *Journal of Value Inquiry*, 35: 449-54.

Ziff, Paul (1960). *Semantic Analysis.* Ithaca: Cornell University Press.

Zimmerman, Michael J. (1983a). "Evaluatively Incomplete States of Affairs." *Philosophical Studies*, 43: 211-24.
_____ (1983b). "Mill and the Consistency of Hedonism." *Philosophia*, 13: 317-35.
_____ (2001). *The Nature of Intrinsic Value*. Lanham: Rowman and Littlefield.

LIBRARY OF ETHICS AND APPLIED PHILOSOPHY

1. O. Lagerspetz: *Trust: The Tacit Demand*. 1998 ISBN 0-7923-4874-5
2. W. van der Burg and T. van Willigenburg (eds.): *Reflective Equilibrium*. Essays in Honour of Robert Heeger. 1998 ISBN 0-7923-5066-9
3. J. G. Murphy: *Character, Liberty, and Law.* Kantian Essays in Theory and Practice. 1998 ISBN 0-7923-5275-0
4. B. C. Postow: *Reasons for Action.* Toward a Normative Theory and Meta-Level Criteria. 1999 ISBN 0-7923-5700-0
5. D. Richter: *Ethics After Anscombe.* Post "Modern Moral Philosophy". 2000 ISBN 0-7923-6093-1
6. G. den Hartogh (ed.): *The Good Life as a Public Good.* 2000 ISBN 0-7923-6167-9
7. T. van den Beld (ed.): *Moral Responsibility and Ontology*. 2000 ISBN 0-7923-6255-1
8. M.J. Almeida (ed.): *Imperceptible Harms and Benefits*. 2000 ISBN 0-7923-6464-3
9. J.A. Corlett: *Responsibility and Punishment, Third edition.* 2005 ISBN 1-4020-4147-0
10. M. Gore Forrester: *Moral Beliefs and Moral Theory*. 2002 ISBN 1-4020-0687-X
11. A.W. Musschenga, W. van Haaften, B. Spiecker and M. Slors (eds.): *Personal and Moral Identity*. 2002 ISBN 1-4020-0764-7
12. C. Wilks: *Emotion, Truth and Meaning*. In Defense of Ayer and Stevenson. 2002 ISBN 1-4020-0916-X
13. M. Schermer: *The Different Faces of Autonomy*. Patient Autonomy in Ethical Theory and Hospital Practice. 2002 ISBN 1-4020-0984-4
14. H. Dyke (ed.): *Time and Ethics: Essays at the Intersection*. 2003 ISBN 1-4020-1312-4
15. J. Ryberg and T. Tännsjö (eds.): *The Repugnant Conclusion*. Essays on Population Ethics. 2004 ISBN 1-4020-2472-X
16. J. Ryberg: *The Ethics of Proportionate Punishment*. A Critical Investigation. 2004 ISBN 1-4020-2553-X