Philosophical Issues, 19, Metaethics, 2009

INTRINSIC VALUES AND REASONS FOR ACTION

Ralph Wedgwood
Merton College, Oxford

0. Introduction

What reasons for action do we have? What explains *why* we have these reasons? In this paper, I shall articulate some of the basic structural features of a theory that would provide answers to these questions. So my primary focus here is on the nature of reasons for action themselves, not on the meaning of the terms that can be used to talk about such reasons. However, it seems plausible that the term 'reason for action' is in fact used in many different ways, and can express many different concepts in different contexts. Hence, some prefatory remarks about the term 'reason for action' are in order, to make it clear which of the many senses of the term I have in mind.

As I am understanding the term here, a "reason for agent x to do act A at time t" is some fact about A (in relation to x's situation at t) that plays a certain sort of role in *explaining* what x *ought* to do at t.[1] The kind of "explaining" that I have in mind entails a sort of *supervenience* thesis: there cannot be a difference between two situations with respect to what the agent ought to do in those situations unless there is also a difference with respect to the reasons that the agent has for and against the various available courses of action in those situations. Unfortunately, however, I cannot undertake a detailed study here of the way in which the agent's reasons explain what the agent ought to do. For the purposes of this paper, I shall just have to rely on an intuitive idea of what it is for these reasons to explain what the relevant agent ought to do.

In fact, however, there is also a further problem with this elucidation of the relevant sense of 'reason for action'. As I have already claimed, it is plausible that terms like 'reason for action' are not univocal, but are instead capable of expressing several different concepts in different contexts. But if

this is true of the term 'reasons for action', then it is surely also true of the term 'ought' as well. Indeed, I have argued in some of my earlier work (Wedgwood 2007, chap. 5) that the term 'ought' is multivocal and context-sensitive in precisely this way. As I propose to use the term 'reason for action' here, the occurrence of the term 'ought' in the elucidation that I have just given of what "reasons for action" are should be read as what I have elsewhere called the *objective practical* 'ought'. (In my view, there are also other kinds of 'ought' as well; and so there are also other kinds of "reasons for action"—where "reasons for action" of each of these other kinds would be facts that contribute towards explaining what the relevant agent "ought" to do, in some *other* sense of 'ought' besides the objective practical 'ought'. But I shall simply ignore all these other kinds of reasons here.)

The distinctive feature of the *practical* 'ought' is that it is the sort of 'ought' that is naturally used to express the conclusion of pieces of practical advice or deliberation about what to do, or to ask deliberative questions about what to do. It seems to me that this sort of 'ought' is neither a narrowly moral 'ought', nor a narrowly prudential 'ought', nor the sort of 'ought' that is relativized to any particular end or goal. Instead, it is simply the general all-things-considered practical 'ought'. Admittedly, some philosophers are sceptical about whether this sort of practical all-things-considered 'ought' really makes sense.[2] In my view, such scepticism is completely unjustified. But unfortunately I shall simply have to assume here that this sort of 'ought' really does exist. This assumption is at least relatively familiar; and so even philosophers who reject this assumption should be interested in seeing how a theory of reasons for action could be developed on the basis of this assumption.

What makes it the *objective* practical 'ought' is that what an agent, in this objective practical sense, ought to do is determined by *all* practically significant facts about the agent's situation—regardless of whether the agent in question actually knows those facts, and even of whether the agent is in a position to know these facts. Of course, one is very often not in a position to know what one—in this objective practical sense—ought to do. But knowing what one—in this objective practical sense—ought to do is the ideally well informed state in which to act; and rational practical reasoning involves making an effort to come as close as the available time and information allow to this ideally well informed state.[3]

This, then, is how I shall be using the term 'reasons for action' in this paper. The goal of this paper is to articulate a general theory of reasons for action—where the term 'reasons for action' is understood in the sense that I have just explained. As I said, my goal is to *articulate* this theory. Although I find the theory attractive, I shall not really have time to argue in support of this theory here. Instead, I shall just present the theory, in a way that will I hope reveal the theory's internal coherence. But a full defence of the theory will have to await another occasion.

The theory that I shall outline in this paper aims to be an account of *all* reasons for action, not just a theory of *moral* reasons for action. I believe that this theory of reasons for action could be used to ground a theory of moral reasons. Roughly, we could say that the distinguishing mark of moral reasons is that it is appropriate to respond to agents' degree of compliance with moral reasons with *reactive attitudes*, such as blame, indignation, or gratitude (whereas if a reason is not a moral reason, then it is not appropriate to respond to an agent's degree of compliance with the reason with reactive attitudes of this sort).[4] But I shall say nothing more about moral reasons here; the focus will be on the nature of reasons in general.

In a nutshell, the distinctive feature of this theory of reasons for action is that it implies that all reasons for action are grounded in facts about how the courses of action available to the relevant agent at the relevant time are related to the *intrinsic values*.[5] As I shall explain, the way in which I conceive of these intrinsic values is in many ways quite similar to the way in which they have been conceived by G. E. Moore (1922) and his followers—such as Michael Zimmerman (2001).

As I have already said, the theory of reasons for action that I shall outline here implies that all reasons for action are generated by intrinsic values. Now, many philosophers assume that the only sort of reason for action that could be generated by these intrinsic values is a reason to *promote* these values, in something like the sense that is familiar from *consequentialist* moral theories. However, my theory of reasons for action is designed to make room for a resolutely *anti-consequentialist* view. Indeed, my theory is designed to be compatible with the *extreme* anti-consequentialist view that *no* reasons for action are simple reasons to "promote" intrinsic values, of the kind that the consequentialists believe in.

How could the intrinsic values generate such non-consequentialist reasons? Some non-consequentialist philosophers have thought that it was so clear that intrinsic values would generate consequentialist reasons for action, if they existed, that they have gone so far as to deny the very existence of such intrinsic values.[6] Clearly, the theory that I shall outline here does not follow these philosophers on this point.

Other philosophers have toyed with the idea that the consequentialists are radically mistaken about the *structure* of values. First, according to some of these philosophers, consequentialists are wrong to focus exclusively on values that are instantiated by *states of affairs*; that is, according to some of these philosophers, it is a crucial fact about these intrinsic values that they are instantiated by items of other kinds, and not just by states of affairs. Secondly, some philosophers reject the central consequentialist idea that these values generate a consistent overall *ranking* of the items that instantiate them to some degree or other. Thirdly, some philosophers reject the common consequentialist idea that intrinsic values have a fundamentally *aggregative*

structure (so that, for example, if it is a good thing for two people's lives to be saved, it is an even better thing for four people's lives to be saved).[7]

According to the theory of reasons for action that I shall outline here, these consequentialist assumptions about the structure of values are essentially correct. Nonetheless, the theory that I shall outline does *not* imply that the only reasons for action that these values generate are consequentialist reasons to promote these values. Admittedly, this theory *could* be developed in some ways that would lead to a consequentialist view of reasons for action; but as I shall explain, it can also be developed in a way that would accommodate a much more anti-consequentialist view.

1. Absolute and Relative Values

In this section, I shall explain in some more detail how I conceive of the intrinsic values that form the basis of this theory of reasons for action.

Since 'good' is the most general positive evaluative term in ordinary English, all values can, broadly speaking, be understood as *ways of being good*.[8] However, within the whole domain of values, we can draw a distinction between what we could call "absolute values" and "relative values".

Some of the ways in which a thing can be good are ways in which it can be good *for* someone or something, in the sense of being *beneficial* for someone or something that can be benefited. Thus, something might be good for you, or good for me, or good for Oxford University, or good for the flowers in your garden, or good for the city's cockroach population. If I say that something x is good for y (for example, that x is good for the Taliban), there need be no reason for you to expect that I am in any way in favour of x: I am simply stating that x benefits y, and if I am bitterly opposed to y, I will in all likelihood not in be in favour of x at all.

A second way in which a thing can be good is a way in which it can be good *for* some end or purpose, in the sense of being *useful* for that end or purpose. Again, if I say that a certain technique is good for inflicting pain on people, you may have no reason to expect me to be in favour of that technique; I am simply stating that that technique is useful for accomplishing the end of inflicting pain. The same point applies with a third way of being good, which is to be good *at* some activity in the sense of being technically *skilled* or *efficient* at that activity. If I say that someone is good at torturing people, you would have no reason to expect me to be in favour of him. The "attributive" use of the term 'good'—as when we say something of the form 'x is a good F', where 'F' is a sortal term like 'knife' or 'dentist'—seems usually to convey that the object in question is good in either this second or third way.[9]

Contrasting with these three kinds of relative goodness, there are also various ways in which things can be *absolutely* good. For example, if I say

that the 1998 Human Rights Act in the UK was a *good thing*, you *will* have a reason to take me to be in favour of the Human Rights Act.

One way in which we can bring out the difference between relative and absolute values is to invoke what I shall call the "fitting attitude equivalence" (or "FA equivalence" for short). According to the FA equivalence, something is good or valuable in a certain way if and only if it is a fitting object of an attitude of a certain corresponding sort. We need not assume that this FA equivalence is an *analysis* in the sense that either the left- or the right-hand side of the biconditional is somehow more fundamental than the other. Still less need we assume what has come to be known—following T. M. Scanlon (1997)—as the "buck-passing" view of value, according to which the FA equivalence shows that it is not a thing's value that *makes* it a fitting object of the corresponding attitude (or that *gives* us a reason for the corresponding attitude), but rather that its value *consists* in its having other features that make it a fitting object of such an attitude. All that we have to assume is that this biconditional connection holds between the value and the fittingness of the corresponding attitude.

If this FA equivalence is correct, then the difference between absolute and relative values can be captured in a fairly simple way. If the property of being F is an absolute value, then an item x has this property if and only if it is appropriate for *anyone* who adequately considers x to have the corresponding attitude; moreover, this attitude will be a pro-attitude of some fairly straightforward kind. If it is really is appropriate for absolutely *anyone* who adequately considers x to have this sort of pro-attitude towards x, this pro-attitude must be an essentially *disinterested* pro-attitude—that is, a pro-attitude that does not depend for its appropriateness on the particular relation that the thinker has towards x. Some central examples of this sort of pro-attitude would be such attitudes as *admiration*, or any other attitude that involves contemplating something with an attitude of disinterested pleasure.

By contrast, if the property of being G is a *relative* value, then this simple form of the FA equivalence does not hold. At most, some more complicated form of the FA equivalence may hold instead, although it is controversial exactly which form of the FA equivalence (if any) will hold for these relative values.

Broadly speaking, there are two possible approaches to giving some sort of FA equivalence for these relative values. According to the first of these two approaches, it is not true that whenever something y exemplifies this relative value G, that makes it fitting for *anyone* to have a pro-attitude towards y; at most, it is fitting for *certain agents* who are somehow related to this value G to have a corresponding pro-attitude towards y. For example, let us focus on the relative value of being *good for Alfred*. Perhaps the most that is true of this relative value is that if y is good for Alfred, then that makes it appropriate for *Alfred* to have a certain corresponding pro-attitude towards y.

A second approach to these relative values would not restrict the *range of agents* for whom the exemplification of the relative value makes a corresponding pro-attitude fitting, but would instead revise its conception of the *sort of attitude* that the value makes fitting. According to this second approach, an item y will have this property of being G if and only if it is appropriate for anyone who adequately considers y to have the corresponding sort of *relative* or *conditional* pro-attitude towards y. For example, if the property of being G is the property of being instrumentally good for a certain end or purpose E, then the relevant "conditional pro-attitude" might be the attitude of *conditionally* favouring y as a means to the end E, conditionally on the assumption that one is going to pursue end E in an effective way. However, we do not need to determine here whether either of these two approaches is correct. The important point for our purposes is just that no *simple* form of the FA equivalence holds for these relative values, as it does for the absolute values.

In a way, then, there is something essentially *agent-neutral* about the absolute values: if some item x instantiates one of these absolute values, that fact makes it appropriate for *all* agents to have the corresponding disinterested pro-attitude, regardless of those agents' identity, or those agents' relationship to x. In this way, these absolute values contrast with more clearly agent-relative values, like the feature of being good or beneficial *for me* or *for you* (where you and I are agents).

I have spoken of "absolute values" here, in the plural (not "absolute value", in the singular), because the theory that I am outlining here is meant to be compatible with a radical *pluralism* about values—that is, with the view that there are many different ways in which things can be absolutely good or valuable, and that there is no way of reducing all these many different ways of being valuable or good to one central master way of being valuable or good.

The idea of an irreducible plurality of values is familiar, but it is in fact an extremely challenging question how best to analyse this idea. However, we can still enumerate a number of respects in which different values seem, at least *prima facie*, to differ from each other. Thus, one way in which different values seem to differ is in being the fitting objects of different sorts of pro-attitude. (Thus, the sort of attitude that is appropriate in response to sublimely beautiful natural phenomena seems importantly different from the sort of attitude that is appropriate in response to admirable intellectual achievements.) Another way in which these absolute values differ is that they are exemplified by different ranges of items; and even when the same items exemplify more than one of these values, the ranking of these items in terms of the degree to which they exemplify one of these values may differ from the ranking in terms of the degree to which they exemplify another one of these values. Unfortunately, a complete investigation of these issues would detain us too long at this point. For our present purposes, the important point is

just that the theory that I shall outline here is designed to be compatible with the idea of an irreducible plurality of values.

2. The Structure of Intrinsic Values

As I have mentioned, my theory of reasons for action is designed to accommodate a resolutely anti-consequentialist view. Nonetheless, my conception of these values is in many respects quite similar to the conceptions that have been developed in the consequentialist tradition. Above all, like many consequentialists, I am happy to restrict my attention entirely to values that are instantiated by *states of affairs*. Moreover, I think of these states of affairs as essentially *abstract* entities, in the sense that a state of affairs can *exist* even if it does not *obtain*. (For example, even if Chris dies at the age of 55, the abstract state of affairs of Chris's living to the age of 85 still exists, even though it is not a state of affairs that actually obtains.)

Some opponents of consequentialism think that one of the basic mistakes that consequentialists make is to concern themselves only with the kind of value that is exemplified by *states of affairs*.[10] But as I shall argue, attacking this focus on the values of states of affairs is a hopeless manoeuvre for the opponents of consequentialism to make. At the very worst, focusing exclusively on states of affairs would be a harmless "housekeeping" move with no theoretical consequences of any importance. (In fact, however, I shall tentatively suggest that there is a positive theoretical advantage in focusing purely on the values of states of affairs—so that this focus on states of affairs should be accepted by all ethical theorists, whether they are consequentialists or not.)

Even though I am going to focus exclusively on the values of states of affairs, it must be conceded straight away that many things other than states of affairs are valuable in various ways. A poem is not a state of affairs, but it can be admirable, and so valuable in one distinctive way; a landscape is not a state of affairs, but it can be sublimely beautiful, and so valuable in another way; an individual human being is not a state of affairs, but individual human persons have a certain sort of dignity that makes each of them valuable in yet another way.

However, whenever something x that is not itself a state of affairs has a given evaluative property V, there is a simple way in which we can identify a corresponding state of affairs $S(x)$ that has a corresponding evaluative property V'. Then instead of talking about the valuable thing x, we can just talk about the correspondingly valuable state of affairs $S(x)$ instead. In other words, the valuable state of affairs $S(x)$ can always serve in our theory as a proxy for the valuable thing x.

This is how to identify the corresponding state of affairs for any valuable thing x (assuming that x is not itself a state of affairs). I assume that whenever

something has a certain evaluative property V, it has some other property P that *makes it the case* that it has V. Then the state of affairs that corresponds to the valuable thing x is the state of affairs of x's having P. This state of affairs may not itself have V, but it has a corresponding sort of value: namely, it is a state of affairs that makes it the case that something has V'—which seems to be a way of being a valuable state of affairs.

Suppose that the state of affairs that instantiates one of these values is identified in the way that I have just outlined, so that it is the state of affairs of x's having property P—where P is the property that makes it the case that x has value V. If P really makes it the case that x has V, then it seems that x's having P must be *sufficient* for x's having V; that is, it is *impossible* for x to have P without also having V. So the state of affairs of x's having P must similarly be sufficient for that state of affairs' having the corresponding value V': that is, it is impossible for this state of affairs to exist without having that value V'. In that sense, this value V' is an intrinsic feature of this state of affairs. So the absolute values instantiated by these states of affairs are intrinsic values—in the sense of philosophers like Moore (1922) and Zimmerman (2001).

If a state of affairs S is intrinsically valuable in this way, then S will have to include a lot more detail than we would normally explicitly indicate when we talk about such valuable states of affairs. Most of the states of affairs that we talk about do not have the feature of its being impossible for them to exist without being valuable. For example, the state of affairs of Chris's living to the age of 85 is not valuable if Chris spends all of his life in a state of excruciating pain and dementia. So the state of affairs of Chris's living to the age of 85 is not intrinsically valuable; it is only a more detailed state of affairs—such as the state of affairs of Chris's living to the age of 85 entirely free from any serious pain or disease—that can be intrinsically valuable. To appropriate Michael Zimmerman's (2001, 142) terminology, states of affairs that are not sufficiently detailed to have any degree of intrinsic value can be called "evaluatively inadequate". The states of affairs that have *more* detail than is necessary in order to determine what degree of intrinsic value they have can be said to have "non-basic intrinsic value". The states of affairs that have exactly the right amount of detail to determine what degree of intrinsic value they have can be said to have "basic intrinsic value".[11]

From now on, I shall simplify our discussion by focusing exclusively on *intrinsically* valuable states of affairs. (There will be no loss of generality in ignoring the *extrinsically* valuable states if it is correct to assume that every extrinsically valuable state of affairs is a part of some larger and more detailed state of affairs that has (either basic or non-basic) intrinsic value.)

Some philosophers—most notably, Zimmerman (2001, Chap. 3)—will object to my focusing on *abstract* states of affairs. According to these philosophers, it is not abstract states of affairs, but *concrete* states that are the fundamental bearers of intrinsic value—where the crucial difference between

abstract states of affairs and concrete states is that unlike abstract states of affairs, concrete states exist only when they actually obtain. The reason for thinking that it is concrete states that are the fundamental bearers of intrinsic value is simply that it seems odd to say that the state of affairs of Chris's living to the age of 85 is intrinsically valuable in some way if as a matter of fact Chris actually dies at the age of 55.

We should certainly agree that ordinarily, any statement of the form 'It is good that *p*' entails that it is *true* that *p*. But it seems to me of great importance to ethical theory to be able to *compare* states of affairs that do not actually obtain; for example, we need to be able to say that the state of affairs of Chris's living to the age of 85 (free from disease and pain and the like) is better than the state of affairs of his dying at the age of 55. For this comparison between two entities to be true, it seems that both entities must exist. So it seems that there is a positive theoretical advantage to focusing on the degrees of value that are instantiated by abstract states of affairs that do not actually obtain.[12]

In what follows, then, I shall concentrate on the values that are instantiated purely by states of affairs of this kind. I shall sometimes allow myself to speak more concisely—for example, by talking about the value of *lives* or *attitudes* or the like. This should be taken as equivalent to talking about the value of the state of affairs that consists in the relevant being's having a life of the relevant kind, or the state of affairs that consists in the relevant thinker's having an attitude of the relevant type on the relevant occasion.

A further respect in which my conception of these intrinsic values is similar to the consequentialists' conception is that I think of these values as fundamentally *coming in degrees*. That is, each of these values is instantiated by some states of affairs to a greater degree than by other states of affairs. As a result, each of these values generates a *ranking* of states of affairs. For example, consider the value that is exemplified by long healthy life. It seems to me that this value is exemplified to an even higher degree by Chris's living a long healthy life for 85 years than by Chris's living a long healthy life for only 75 years instead.

This is not to say that every single value ranks *all* states of affairs whatsoever. On the contrary, it may be that each of these many values is instantiated only by a restricted range of states of affairs. And even when the same value is instantiated by two different states of affairs, we still cannot assume that the degree to which the value is instantiated by the first state of affairs is either greater or less than, or equal to, the degree to which it is instantiated by the second. The relevant value may only provide a *partial* ordering of these states of affairs.[13] However, there are also many cases in which two states of affairs *are* ranked by one of these values. One particularly common case in which a pair of states of affairs are ranked by one of these values is when those states of affairs are broadly similar to each

other, but are also incompatible with each other (that is, it is impossible for more than one of those states of affairs to obtain). Thus, in the example that I have just given, the state of affairs of Chris's living a long healthy life for 85 years and the state of affairs of his living a long healthy life for just 75 years are broadly similar, but mutually incompatible, states of affairs.

Indeed, it may well be that these *rankings* of states of affairs are actually more fundamental than any non-comparative fact about how a particular state of affairs instantiates one of these values. That is, it may be that John Broome's (1999b) thesis that "Goodness is reducible to betterness" is correct. If this thesis is correct, then there is no non-arbitrary dividing line between good and bad, or between value and disvalue; there is only better and worse, more and less valuable.[14] When we make a non-comparative statement, of the form 'The state of affairs S_1 is bad', our statement is true in virtue of the fact that the state of affairs S_1 is at least somewhat worse than the *contextually relevant standard* (which is very commonly something like the state of affairs that might have been *expected* in the context of the statement). I shall not take any definite stand here about whether this thesis is correct (although as a matter of fact, I am sympathetic to this thesis); but I shall try to make everything that I say compatible with this thesis.

One final way in which my conception of values resembles that of traditional consequentialism is that I am inclined to think that these values typically have a fundamentally *aggregative* structure. For example, if there are ten people who are in danger of drowning, then it is better for five of those people's lives to be saved than for just one of their lives to be saved, and it is even better for all ten of their lives to be saved than for just five of their lives to be saved. Similarly, if it is a great achievement to prove one important mathematical theorem, it is an even greater achievement to prove two such theorems. If it is a bad thing for one species of butterfly to go extinct, it is even worse for two such species to go extinct, and so on. It is a challenging question how to give a precise general formulation of the sort of aggregative structure in question.[15] But we need not pursue this question in detail here. Again, I need not take a definite stand on whether the intrinsic values really do have this aggregative structure, nor on what exactly this aggregative structure comes to. I shall simply try to make everything that I say compatible with the assumption that the intrinsic values are indeed aggregative in something like this way.

3. A Sketch of a Substantive View of Values

It may help to make the theory that I am articulating here easier to understand if I give a rough sketch of a substantive view of these intrinsic values. I need not presuppose all the details of this precise view of values.

Instead, this section gives a sketch of a substantive view of values purely for the sake of illustration.

In giving my sketch of this view of the intrinsic values, I shall deploy a distinction between the *simple* intrinsic values, and the *complex* intrinsic values. We can explain this distinction in the following way. A complex intrinsic value can only be instantiated by a complex state of affairs that itself involves the instantiation of some value by some *other* state of affairs.[16] By contrast, the simple intrinsic values can be instantiated by simple states of affairs, which do not themselves involve the instantiation of any other value by any other state of affairs.

Broadly speaking, the view that I shall sketch in this section is a generalization of some of the ideas of the *perfectionist* tradition in ethics—as exemplified by the ethical theories such philosophers as Aristotle and Aquinas, and in more recent times by the theories of T. H. Green and Thomas Hurka (1992). Whereas traditional perfectionism focused on perfections of *human nature*, a more general theory could focus on the perfections of *living nature*.

One form that these "perfections of living nature" can take is when living things, such as plants and animals (including both human and non-human animals), live long healthy lives, free both from pain and from disease. As I explained in the previous section, however, I mean this theory to be compatible with Broome's thesis that "goodness is reducible to betterness". So, we may take this value to consist most fundamentally in the way in which it is *more* valuable (other things equal), with respect to this value, when living things live longer, or have better health, or less pain or disease.

More controversially, I suggest that it is another feature of this value that it is also more valuable (other things equal), with respect to this value, when *more* living things exist. (So, for example, according to this view, the *longer* the period of time during which the universe supports life, the more valuable the state of affairs is, with respect to this value.) It may be yet another feature of this value that (other things equal) it is more valuable (with respect to this value) when a greater *diversity* of living things exist. So, if this view is correct, then (other things equal) greater species diversity is in this respect more valuable than less.

Other simple values might be perfections of more distinctively *human* life. The perfections of *experience* might include discerning perceptions of beauty and other valuable aesthetic qualities. Perfections of *cognition* and other distinctively human capacities might include the development and exercise of skills and talents of various kinds (including intellectual, artistic, and athletic achievements).

In addition to these simple values, however, there are also *complex* values. As I have explained, the complex intrinsic values can only be instantiated by a complex state of affairs that itself involves the instantiation of some intrinsic value by some *other* state of affairs. One central instance of a complex value,

as Hurka (2001, 13–14) has argued, is the value of *loving* and *admiring* things of intrinsic value. It is itself something of intrinsic value, I suggest, when intelligent beings love and admire things of intrinsic value, in the way in which those valuable things merit. Similarly, it is of intrinsic value for intelligent beings to hate and deplore things that merit being hated and deplored.

The value of loving things of intrinsic value may be related the value of *well-being*, at least if the correct theory of well-being is something like the theory that has been defended by R. M. Adams (1999). According to Adams, what is fundamentally good for a person—and so constitutes that person's well-being—is a life that is characterized by *enjoying* things of intrinsic value. If loving and admiring things of intrinsic value is itself of intrinsic value, it seems plausible that it will also be of intrinsic value for people to enjoy things of intrinsic value—and according to a view like Adams's, a life characterized by such enjoyments is precisely what well-being consists in.[17]

A third important complex value is the value of *interpersonal relationships* (such as friendships) that advance the well-being of at least some of the people involved in the relationship, or in some other way help some of those who are involved in the relationship to succeed in some valuable or worthwhile activity. Besides the value of the existence of these relationships, there is also the value of the *attitudes* and *activities* that are distinctive of these relationships: these distinctive attitudes include the special concern that we typically have for our friends' well-being, and for the success of their projects and activities; and the distinctive activities of friendship include above all co-operative activities of collaborating to achieve some shared goal, and the sort of conversation that involves mutual communication and sharing of ideas, experiences, and so on.

For our purposes, however, the most important complex values are the values exemplified by *courses of action*. (I shall use the phrase 'course of action' broadly, so that it includes *omissions* as well as *acts*.) I shall discuss this topic in the next section.

4. The Intrinsic Values of Courses of Action

Strictly speaking, since we are focusing only on the values that are exemplified by states of affairs, my talk about the values that are exemplified by a course of action should be taken to refer to the values that are exemplified by the *state of affairs* that consists in the relevant agent's taking that course of action at the relevant time. Moreover, if this state of affairs is to exemplify a complex *intrinsic* value, the relevant "course of action" must be a rather more specific and detailed type of behaviour than the ones that we normally talk about. For example the relevant course of action might be something like: acting in such a way that a state of affairs that instantiates

value V to degree d_1 results, when an alternative way of acting would have resulted in a state of affairs that instantiates value V to degree d_2 instead; or something of that general kind.

According to the view that I shall propose in this section, there are two main components to the absolute value of a course of action. First, there are various *states of affairs* each of which counts as a *consequence* of the course of action, in the sense (roughly) that all these states of affairs will obtain if the course of action is undertaken. For example, it might be that one state of affairs that counts as a consequence of a certain course of action A_1 is the state of affairs—call it $S_{1.1}$—of five people in the Library being saved; and another consequence of the same course of action A_1 is the state of affairs $S_{1.2}$ of one person in the Common Room being killed. (Unlike consequentialism, the theory that I am sketching here does not just aggregate all of the consequences of A_1 into one big consequence. On the contrary, it will often be important to disaggregate these consequences, treating the course of action A_1 as having *several* different consequences.) So, one component of the value of this course of action A_1 consists of the intrinsic values that are instantiated by each of the states of affairs—such as $S_{1.1}$ and $S_{1.2}$—that count as its consequences.

If this approach is to be compatible with Broome's thesis that "goodness is reducible to betterness", then the value of each of the consequences of the course of action must be measured *comparatively*. Somehow, then, for every agent and every time when the agent is capable of acting, and for every intrinsic value V that is exemplified by some of the consequences of courses of action that are available to that agent at that time, there is a relevant "benchmark of comparison" for V with respect to the situation of the agent at that time. When a consequence of a course of action is *superior* to the benchmark (with respect to the relevant value), the consequence exemplifies this value to a *positive* degree: in other words, having this consequence counts as a *good* feature of the course of action. When a consequence of a course of action is *inferior* to the benchmark (with respect to that value), the consequence exemplifies this value to a *negative* degree; and that counts as a *bad* feature of the course of action.[18]

It will require much further investigation to understand what determines where this relevant benchmark of comparison lies. But here is one suggestion that may help to fix ideas. Consider the set of alternative courses of action that are available to the agent at the relevant time; let us say that the "relevant alternative options" are all the alternative courses of action that might be considered seriously by a rational deliberator. Suppose that each of these "relevant alternative options" A_i has a consequence S_i that exemplifies the value V to some degree—such that collectively these consequences form a set of states of affairs each member of which is incompatible with every other. Call these consequences "the relevant alternative consequences." Perhaps the relevant benchmark of comparison is simply the *average*

degree to which these relevant alternative consequences exemplify this value V.

For example, suppose that in the situation of the agent at the relevant time, there is just one alternative course of action A_2—besides A_1—that might be seriously considered by a rational deliberator; A_2 has two significant consequences, $S_{2.1}$ and $S_{2.2}$—where $S_{2.1}$ is the state of affairs of the five people in the Library being killed, and $S_{2.2}$ is the state of affairs of the one person in the Common Room being saved. Let us focus on how these consequences compare in terms of the value of well-being. The relevant comparisons will be between the "relevant alternative consequences"—in this case, between $S_{1.1}$ and $S_{2.1}$, and between $S_{1.2}$ and $S_{2.2}$. If the benchmark for each of these two comparisons is the average value of the two consequences that are being compared, then clearly $S_{2.2}$ (the one person in the Common Room being saved) is superior to the benchmark; and so it counts as a *good* feature of A_2 that it has this state of affairs as a consequence. On the other hand, the alternative consequence $S_{1.2}$ (the person in the Common Room being killed) is inferior to the benchmark; so it counts as a *bad* feature of A_1 that it has this state of affairs as a consequence. (Obviously, the position of A_1 and A_2 is the other way round with respect to the comparison between the two consequences that concern the five people in the Library.)

However, besides the values of each of the consequences of a course of action, I propose that there is also a *second* component to the value of a course of action. This second component is what I shall call the agent's *degree of agential involvement* in bringing about each of those consequences.

There is, it seems to me, a whole gamut of degrees to which one may be agentially involved in bringing it about that one state of affairs S_1 obtains rather than an alternative state of affairs S_2. This notion of an agent's degree of agential involvement in bringing about a state of affairs clearly needs to be studied in depth, and I shall only be able to make a few tentative remarks here.

Broadly speaking, I suggest that there are two dimensions along which one can be to a greater or lesser degree agentially involved in bringing about a state of affairs: the first dimension is *causal*; the second dimension is *intentional*.

Along the causal dimension, there is a particularly crucial difference between *actively causing* a state of affairs and merely *failing to prevent* that state of affairs (in effect, this is what many philosophers think of as the distinction between *doing* and *allowing*). If you merely *fail to prevent* a state of affairs from coming about, then your degree of agential involvement in bringing about that state of affairs is much less than if you *actively cause* that state of affairs to come about. However, even if you actively cause a state of affairs to come about, it still seems to me that there is a wide difference between different cases, depending on the degree to which you are agentially

involved in bringing about this state of affairs. We can illustrate this point by considering two versions of the notorious Trolley Case.[19]

In the original Trolley Case, a runaway trolley is hurtling along a railway towards five people who are trapped on the track. In this case, you can save the five by pulling a lever, which will divert the trolley away onto a side track. Unfortunately, there is a person who is trapped on the side track, and you foresee that if you divert the trolley, the trolley will hit him and kill him. By contrast, consider Frances Kamm's (1996, 173) Grenade Case: if you throw a grenade at the trolley, it will cause the trolley to blow up before it can reach or harm the five, but the shrapnel from the blast will kill an innocent bystander.

In both cases, you actively cause the death of one person; you do not merely fail to prevent his dying from other causes (and in both cases, it is no part of what you *intend* that he should die, or that he should be hit by the trolley or by the shrapnel from the blast). However it seems to me that in the Grenade Case you are agentially involved in causing the death of the bystander to a greater degree than you are in causing the death of the one in the original Trolley Case. In original Trolley Case, a lethal process is already under way, and you merely deflect this process from one trajectory onto another on which it will cause less harm; this is the only causal role that you play in causing the death of the one. By contrast, in the Grenade Case, you initiate a new threat (the grenade) that did not exist independently of your action, and the bystander dies as a result of that new threat. So, in the Grenade case, your degree of agential involvement in bringing about the death of the bystander seems greater than your degree of agential involvement in bringing about the death of the one in the original Trolley case.

In addition to this causal dimension of agential involvement, there is also the *intentional* dimension. Other things equal, your degree of agential involvement in bringing about a state of affairs is greater if you directly *intend* that state of affairs than if you merely *foresee* that that state of affairs will result from your action. (It may also make a difference whether the state of affairs is a highly *important* part of your intention, or whether it is a relatively *incidental* part of your intentions.) Your degree of agential involvement in bringing about a state of affairs may also reflect the amount of *thought* and *effort* that you had to put into bringing about that state of affairs. The underlying idea here is that the greater the amount of thought and effort that you have to put into bringing about a bad state of affairs, the greater the degree to which as Thomas Nagel (1986, 181) put it, your "will" is being "guided by evil".

We can illustrate this difference by contrasting the original Trolley case with a variant of this case, which I shall call the Multiple Loop case.[20] In this case, there is an immense spaghetti junction of tracks, and so there are numerous ways in which you can divert the trolley off the main track; but

unfortunately, unless the trolley smashes into one person who is trapped on one of the side tracks, it will loop round and hit the five from the other side. In fact, the only way to get the trolley *not* to loop back towards the five is by deftly manipulating the trolley through this complex series of junctions, in just such a way that it hits the one person and (as a result of the impact) grinds to halt before it can loop round to hit the five. As I intend to use the term, your degree of "agential involvement" in bringing about the death of the one is much greater in the Multiple Loop case than in the original Trolley case.

No doubt this idea of the "degrees of one's agential involvement" in bringing about a given state of affairs needs a lot more clarification. But in general, it seems plausible that if one of the consequences of a course of action is worse than the relevant benchmark, then the contribution that the badness of this consequence makes to the badness of the course of action is *magnified* by the degree of one's agential involvement in bringing about that consequence. It seems also true to me that if a consequence of a course of action is *better* than the relevant benchmark, then this *good* feature of the course of action is also *magnified* by the degree of the agent's agential involvement in bringing about that consequence.[21]

For example, suppose that one of the states of affairs that results from a certain course of action that is open to you is that Polyxena dies, many decades earlier than she would have done had you acted otherwise. This is a bad feature of the course of action. However, the degree of badness varies considerably with your agential involvement in bringing about Polyxena's death. The more agentially involved you are in bringing about her death, the worse your course of action is; the less agentially involved you are, the less grave a feature of your course of action it is that its consequences include her premature death.

This then is my proposal about the intrinsic values that are exemplified by courses of action. The main components in explaining the goodness or badness of courses of action are twofold: first, there is the relative value of each of the various consequences of the course of action, measured against the relevant benchmark of comparison; and secondly, there is the agent's degree of agential involvement in bringing about each of those consequences.

5. Reasons and Values

With the account that I have just given of the complex intrinsic values of courses of action, I am now in a position to articulate the core of the theory of reasons for action that I am outlining here. Roughly, the theory says that a reason in favour of a course of action is simply an *intrinsically good feature* of that course of action. In other words, x has a reason

to take course of action A at t if and only if the state of affairs of x's doing A at t itself *instantiates* or *exemplifies* one of these complex intrinsic values.

We can put this theory of reasons for action together with our account of the complex intrinsic values of courses of action, in the following way. Consider a *weighting* of the degree to which a consequence S of a course of action A instantiates an intrinsic value V by the agent's degree of agential involvement in bringing about S. (The higher the degree of agential involvement, the greater the weighting; the lower the degree of agential involvement, the lesser the weighting.) Call this the *agentially-weighted value* of S (considered as a consequence of A with respect to V).

Then we can express this theory of reasons for action in the following way: Whenever there is a reason in favour of a course of action A, this reason arises from A's having a consequence S that has a *positive* agentially-weighted value (of this kind); whenever there is a reason *against* A, the reason arises from A's having a consequence S that has a *negative* agentially-weighted value.

Moreover, there is a natural way of using this idea to give at least the beginnings of an account of what it is for one reason to count as a *stronger* reason than another. Other things equal, the *higher* the positive agentially-weighted value, the stronger this reason for A is; the *lower* the negative agentially-weighted value, the stronger the reason *against* A. (The qualification 'other things equal' is important here, because there will be many cases of *interactions* between reasons: for example, there may be cases in which the presence of one reason in the situation of an agent may significantly strengthen or weaken another reason; indeed in some cases, one reason may completely silence or cancel out another reason. But unfortunately I cannot attempt to study these phenomena here.)

In the remainder of this paper, I shall explain how this theory of reasons for action (together with the theory about the kind of values that are instantiated by courses of action) is equipped to make room for a hard-line anti-consequentialist view. (In this section, I shall use the term 'consequentialism' exclusively for what is often called *act* consequentialism or *direct* consequentialism.)

As Krister Bykvist (2002) has persuasively put it, the key idea of consequentialism is that the normative status of any action depends entirely on the *value* of the *outcome* of each of the actions available to the relevant agent at the relevant time. By the "outcome" of an action, Bykvist means what we could call the "total outcome"—where the total outcome of an action must include *everything* that will be the case if the action is performed. Indeed, for consequentialists as Bykvist understands them, it would be a completely harmless simplifying move to assume that there is a unique *possible world* that will obtain if the action is done, and to identify the "outcome" of the action with this possible world.

By the "value" of the action's outcome, Bykvist means what I have here called intrinsic value. My theory does not differ from consequentialism in this respect: my theory also views the status of being an action that there is a reason to do—which is presumably a certain "normative status" as Bykvist puts it—as depending on facts about the instantiation of such intrinsic values. My theory differs from consequentialism because it does not view reasons for action as depending on the value of the *total outcome* of each of the available actions; it views reasons for action as depending on the value of the *actions themselves*. On some views about the value of actions, this would be a distinction without a difference. Specifically, if the value of an action itself depended completely on the value of its total outcome, there would be no real difference between a theory that appealed to the value of the outcome and a theory that appealed to the value of the action itself. However, my theory is designed to allow the value of an action to depend on more than just the value of its total outcome. It allows for this because (i) the theory disaggregates the consequences of the action, allowing an action to have several different consequences, and (ii) measures the value of the action by weighting the value of each of its consequences by the agent's degree of agential involvement in bringing about that consequence.

So, suppose that there are two actions available to you, A_1 and A_2. Suppose that the outcomes of A_1 include $S_{1.1}$—you destroy something of great value—and $S_{1.2}$—Fred, a total stranger to you, fails to prevent you from committing this act of destruction. Suppose that the outcomes of A_2 include $S_{2.1}$—*Fred* destroys this object of great value—and $S_{2.2}$—*you* fail to prevent *Fred* from committing this act of destruction. Otherwise, let us suppose, the outcomes of A_1 and A_2 do not differ in any significant way. Since intrinsic values are agent-neutral, the total outcomes of the two actions must have exactly the same intrinsic value. So for the consequentialist, the two actions have exactly the same normative status; there is no reason for preferring either of these of two actions A_1 and A_2 over the other.

For my theory on the other hand, there is likely to be a reason in *favour* of A_2, and a reason *against* A_1. Your degree of agential involvement in the destruction of the valuable object is much greater if you do A_1 and actively cause the destruction of the valuable object yourself, than if you do A_2 and merely fail to prevent its destruction. So the intrinsic badness of the consequence is multiplied by the greater degree of agential involvement in the case of A_1 and not in the case of A_2. So even though the total outcomes have the same intrinsic value, this theory allows that A_1 is a *worse* course of action than A_2, and so can also allow that you have a reason for preferring A_2 over A_1.

Here is another way to bring out the difference between my theory of reasons for action and a consequentialist view. Suppose that you accepted a consequentialist theory of reasons for action; and suppose that you intend

to do what you have reason to do. Then you would in effect be pursuing a completely *impersonal* agent-neutral aim—specifically, the aim of the relevant value's being instantiated in the world as a whole. This is a classic feature of consequentialist theories: the intention to conform to these theories is in effect the intention to pursue an agent-neutral aim; that is the sense in which, as Derek Parfit (1984, 27) put it, a consequentialist theory gives all agents the very same aim.

On the other hand, suppose that you believed the theory that I have just outlined; and suppose that you intend to do what you have reason to do. Then it may not be true to say that you would in effect be pursuing an impersonal or agent-neutral aim. On the contrary, it may be that the only way in which we could describe your aim would be by saying that you were pursuing an aim that was intensely *agent-relative* and indeed *time-relative* as well—specifically, the aim that *your* conduct, *at the present time*, should be good and fine and valuable in the appropriate way. This would not be an impersonal aim concerned with the state of the world as a whole. It would be an aim that focuses on a very particular part of the world—namely, *what you are doing right now*. This is a different aim for each person, and indeed for each time. So this theory (unlike a consequentialist theory) need not give all agents the very same aim.

In this way, this theory of reasons for action aims to accommodate a view that agrees with Bernard Williams's (1985, 84) complaint that consequentialism makes the mistake of conceiving of the agent who acts rightly as the "World Agent"—that is, an agent who acts at all times as if she were constantly pursuing the single supreme end of optimizing the total state of the world, with no special concern for her own part in the world. The theory of reasons for action that I have outlined here is compatible with a quite different picture of the agent who complies with the reasons for action that she has—a picture of an agent who acts, not like the World Agent, but rather like someone who has a special concern for what she is doing, and for what role she is playing in the world around her.

This then is how this theory of reasons for action can accommodate a resolutely anti-consequentialist view of reasons. It can do this even though it views all reasons for action as generated by the relations between the available courses of action and the intrinsic values, and even though it conceives of these intrinsic values in a fundamentally similar way to the way they have been conceived by many consequentialists. According to this theory, a reason in favour of a course of action arises, not from the intrinsic value of the total outcome of the action, but from the values that are instantiated by that course of action itself. This will lead to an anti-consequentialist view if the value of a course of action depends, not just on the values of its various consequences, but also on the agent's degree of agential involvement in bringing about each of those consequences.[22]

Notes

1. This is obviously closely related to John Broome's (2004) definition of a "reason to φ" as a fact that plays the "for-φ role" in a "weighing explanation" of whether or not the relevant agent ought to φ. (According to the main rival approach, a "reason to φ" is a fact that in some way serves as a *starting point for a possible process of practical reasoning* that would lead the relevant agent to be *motivated* to φ; for a clear statement of this rival approach, see Setiya (2007, 12). My own view is that these are simply two different kinds of concept expressed by the term 'reason', and each of these two approaches captures one of these two kinds. But I cannot defend this view here.)
2. For such scepticism, see e.g. Foot (1978, 169–70) and Finlay (2008).
3. This is a very rough and incomplete characterization of what practical reasoning involves; unfortunately, I cannot undertake a full discussion of rational practical reasoning here.
4. Obviously this approach to moral reasons has a long history behind it; see for example the role that such reactive attitudes play in Mill (1871) and Gibbard (1990), among many others.
5. In this way, this theory is most closely akin to the views of Joseph Raz (1999), and of other contemporary philosophers who have sought inspiration in the views of practical reason that were developed in classical antiquity. But unfortunately I shall not be able to undertake any explicit comparison of this theory with either Raz's views or with the views of the ancient Greek philosophers.
6. For some philosophers who seem to take this approach, see Foot (2002) and Thomson (2001 and 1997).
7. For such claims about how consequentialists misunderstand the structure of intrinsic values, see e.g. T. M. Scanlon (1997) and R. M. Adams (1999).
8. For an illuminating discussion of some "ways of being good", see Thomson (2001)—which in turn draws extensively on the work of von Wright (1963).
9. For this reason, I am inclined to agree with Judith Thomson's (1997, 278) scepticism about P. T. Geach's (1967) view that 'good' always functions as a "logically attributive" term.
10. E.g., there are remarks to this effect in T. M. Scanlon (1997); in R. M. Adams (1999); in Philippa Foot (2003, Chapter 4); and in Bernard Williams's contribution to Smart and Williams (1973).
11. I should note that I am not committing myself to Zimmerman's (2001, 156–8) analysis of what it is for a state of affairs to have "basic intrinsic value" (even though I am using the term in what seems to be at least roughly the same sense as he is).
12. Some philosophers will object that these comparisons between states of affairs that do not actually obtain are really comparisons between the *degrees of value* that these states of affairs *would* have if they did obtain. But surely the whole idea of "degrees of value" is less fundamental than the more basic idea of comparisons between items that are better or worse than each other. So this objection does not seem to me to offer a plausible alternative to the view that is defended here.
13. For a valuable discussion of incommensurability and such partial orderings, see Broome (1999a).

14. If Broome's thesis is correct, then we will in fact have to revise the FA equivalence so that it is fundamentally an equivalence about when it is fitting to have certain essentially *comparative* attitudes—such as certain kinds of *preference*, and the like. But I cannot take the time to revise the FA equivalence in this way here.
15. E.g. Zimmerman (2001, Chap. 5) makes the following proposal: if a state of affairs S_1 is the conjunction of two simpler states of affairs S_2 and S_3, and each of S_2 and S_3 has "basic intrinsic value" of a certain kind V, then the degree to which S_1 exemplifies this value V is simply the *sum* of the degrees to which S_2 and S_3 exemplify V. But I cannot undertake to evaluate this proposal here.
16. This notion of a "complex intrinsic value" is clearly akin to Hurka's (2001, 19) idea of "higher-level intrinsic goods". Unfortunately, I cannot take the time here to analyse the precise relationship between my distinction and Hurka's.
17. If this is correct, then whenever a state of affairs is (e.g.) *intrinsically good for Wolfgang* (which is a way of being *relatively* good), then the state of affairs will also exemplify a certain *absolute* value (viz., the intrinsic value of well-being). This does not undermine the distinction that I drew earlier between relative and absolute values: the relative value (*being good for Wolfgang*) and the absolute value (*being good for some person or persons*) will still generate quite different rankings of the relevant states of affairs. Moreover, the simple form of the FA equivalence will still fail for the relative value (*being good for Wolfgang*)—the fittingness of a disinterested pro-attitude towards x has nothing special to do with Wolfgang, and so will not be sufficient for x to be good for Wolfgang.
18. So this benchmark is in effect a sort of zero-point. But the idea of such a benchmark is still compatible with Broome's thesis that "goodness is reducible to betterness" because there is a *separate* benchmark for every practical situation of an agent at a time. There is still no unique zero-point that divides *all* states of affairs that instantiate the relevant value (including states of affairs are not practically available to the same agent at the same time).
19. The Trolley case was originally due to Philippa Foot (1978, 23), although it has since been discussed by an enormous number of moral philosophers.
20. Michael Otsuka presented the Multiple Loop case in a discussion of Frances Kamm's work at the Pacific Division Meeting of the APA in San Francisco in April 2007. In the end, this case was not discussed in Otsuka's (2008) published version of the comments, although it is discussed in Kamm's (2008) response.
21. It is clearly a more admirable achievement to have a higher degree of agential involvement in bringing about a beautiful painting or the proving of an important mathematical theorem, for example; and it seems to me that this generates a reason for playing this sort of active role in producing the painting or in proving the theorem. Interestingly, however, this phenomenon does not seem to be found in life-and-death cases (such as Trolley cases). As I hope to explain elsewhere, this seems to be because in these cases some additional—and distinctively *moral*—reasons come into play, which masks this effect.
22. Earlier versions of this paper were presented at the Université de Montréal, the University of Edinburgh, Ben-Gurion University, and a conference on Reason and Value at the University of California, Santa Barbara. I am indebted to those audiences (and especially to my commentator in Santa Barbara, Ulrike Heuer), and also to R. M. Adams, John Broome, Krister Bykvist, David Charles,

Roger Crisp, Richard Holton, Rae Langton, Michael Otsuka, Derek Parfit, and Timothy Williamson, for many helpful comments on this material.

References

Adams, R. M. (1999). *Finite and Infinite Goods* (Oxford: Clarendon Press).
Broome, John (1999a). "Incommensurable values", in John Broome, *Ethics Out of Economics* (Cambridge: Cambridge University Press): 145–61.
——— (1999b). "Goodness is reducible to betterness: the evil of death is the value of life", in John Broome, *Ethics out of Economics* (Cambridge: Cambridge University Press): 162–73.
——— (2004). "Reasons", in R. J. Wallace, Michael Smith, Samuel Scheffler and Philip Pettit, eds., *Reason and Value: Essays on the Moral Philosophy of Joseph Raz* (Oxford: Oxford University Press).
Bykvist, Krister (2002). "Alternative Actions and the Spirit of Consequentialism", *Philosophical Studies* 107: 45–68.
Finlay, Stephen (2008). "Oughts and Ends", *Philosophical Studies*. DOI 10.1007/s11098-008-9202-8
Foot, Philippa (1978). *Virtues and Vices* (Oxford: Blackwell).
——— (2002). *Moral Dilemmas* (Oxford: Clarendon Press).
Geach, P. T. (1967). "Good and Evil", in Philippa Foot, ed., *Theories of Ethics* (Oxford: Oxford University Press).
Gibbard, Allan (1990). *Wise Choices, Apt Feelings* (Cambridge, Massachusetts: Harvard University Press).
Hurka, Thomas (1993). *Perfectionism* (New York: Oxford University Press).
——— (2001). *Virtue, Vice, and Value* (New York: Oxford University Press).
Kamm, F. M. (1996). *Morality, mortality*, Vol. II (Oxford: Oxford University Press).
——— (2008). "Responses to Commentators on *Intricate Ethics*", *Utilitas* 20: 111–42.
Mill, J. S. (1871). *Utilitarianism*, 4th edition (London: Longmans, Green, Reader, and Dyer).
Moore, G. E. (1922). "The Conception of Intrinsic Value", in G. E. Moore, *Philosophical Studies* (London: Routledge & Kegan Paul): 253–75.
Nagel, Thomas (1986). *The View from Nowhere* (Oxford: Clarendon Press).
Otsuka, Michael (2008). "Double Effect, Triple Effect and the Trolley Problem: Squaring the Circle in Looping Cases", *Utilitas* 20: 92–110.
Parfit, Derek (1984). *Reasons and Persons* (Oxford: Oxford University Press).
Raz, Joseph (1999). *Engaging Reason* (Oxford: Clarendon Press).
Scanlon, T. M. (1997). *What We Owe to Each Other* (Cambridge, Massachusetts: Harvard University Press).
Setiya, Kieran (2007). *Reasons without Rationalism* (Princeton, NJ: Princeton University Press).
Thomson, Judith Jarvis (1997). "The Right and the Good", *Journal of Philosophy* 94 (no. 6, June): 273–98.
——— (2001). *Goodness and Advice* (Princeton, New Jersey: Princeton University Press).
Von Wright, G. H. (1963). *The Varieties of Goodness* (London: Routledge).
Wedgwood, Ralph (2007). *The Nature of Normativity* (Oxford: Clarendon Press).
Williams, Bernard (1985). *Ethics and the Limits of Philosophy* (Cambridge, Massachusetts: Harvard University Press).
Zimmerman, Michael J. (2001). *The Nature of Intrinsic Value* (Lanham, MD: Rowman and Littlefield).