Oxford Studies in Metaethics

VOLUME 3

Edited by
RUSS SHAFER-LANDAU

OXFORD UNIVERSITY PRESS · OXFORD

2

Fitting Attitudes and Welfare

Chris Heathwood

The purpose of this paper is to present a new argument against so-called fitting-attitude analyses of intrinsic value, according to which, roughly, for something to be intrinsically good is for there to be reasons to want it for its own sake. The argument is indirect. First, I submit that advocates of a fitting-attitude analysis of value should, for the sake of theoretical unity, also endorse a fitting-attitude analysis of a closely related but distinct concept: the concept of intrinsic value *for a person*, that is, the concept of *welfare*. Then I argue directly against fitting-attitude analyses of welfare. This argument, which is the focus of the paper, is based on the idea that whereas whether an event is good or bad for a person doesn't change over time, the attitudes there is reason to have towards such an event can change over time. Therefore, one cannot explain the former in terms of the latter, as fitting-attitude analyses of welfare attempt to do.

1. FITTING-ATTITUDE ANALYSES OF VALUE

1.1. Background

G. E. Moore famously argues (1903a, § 13) that the property of being intrinsically good is unanalyzable and that the predicate 'is intrinsically

Versions of this paper were presented at the Center for Values and Social Policy at the University of Colorado at Boulder in April 2006, at the Third Annual Metaethics Workshop at the University of Madison–Wisconsin in September 2006, and at the Pacific Division Meeting of the American Philosophical Association in San Francisco in April 2007. Thanks to those in attendance for helpful discussion, especially Gunnar Björnsson, Tom Carson, Steve Finlay, Pat Greenspan, Christian Lee, Don Loeb, Kris McDaniel, Bob Pasnau, Doug Portmore, Geoff Sayre-McCord, Mark Schroeder, Caj Strandberg, and Michael Tooley. Special thanks to my APA commentator, Bana Bashour. Ben Bradley, Fred Feldman, Jens Johansson, Mark LeBar, Brad Skow, and Stephan Torre each read a version of the paper and provided comments that helped me improve it; I am

good' is indefinable. If he is right, then friends of intrinsic value are in a bit of a bind: we believe in some property—indeed, we hold it to be a very important one, one significantly impacting how we are required to behave—but we've never seen it, we can't tell you what its nature is, and we can't define the word we use to express it. Wouldn't it be better if we could do this?

Some friends of intrinsic value have tried to do this, and in a way against which Moore's arguments lose some of their force.[1] Moore's open question argument is most forceful against naturalistic analyses of value, but these philosophers have proposed a certain kind of not-necessarily-naturalistic analysis: the *fitting-attitude analysis of value*. According to this view, to say that something is intrinsically good is to say, roughly, that we should desire it, or that we have reason to desire it, or to favor it, or to have some other pro-attitude towards it. The theory explains *value*, an axiological notion, in terms of some deontic notion, such as *obligation, fittingness*, or *reasons*. But it is "less reductive" than naturalistic analyses, since it doesn't attempt to reduce away the normativity of value. The analysans of a fitting-attitude theory contains a normative notion.[2]

Fitting-attitude theories are metaethical theories of *what it is* for something to be good, not normative theories of what things are good; so they should be compatible with any theory of the latter. When hedonists tell us that pleasure and pleasure alone is good, what the hedonists are saying, according to fitting-attitude theorists, is that pleasure and pleasure alone is fit to be desired. Fitting-attitude theories are, in a sense, *response-dependent* theories, but the relevant responses are not the ones we do or would have, but the ones we *should* have.

The earliest prominent advocate of this theory is usually thought to be Franz Brentano (1889). But Sidgwick endorses a version of the view in at least the third edition of *The Methods of Ethics* (1884: 108): "I

grateful to each of them. Finally, I'd like to thank those who contributed to a discussion of some of these topics on the weblog PEA Soup.

[1] Not all such philosophers are in fact friends of intrinsic value (if by 'intrinsic value' we mean value that supervenes on intrinsic nature). Some are more interested in theorizing instead about *final value*—the value that something has as an end, a value which it may turn out does not supervene on intrinsic nature (for discussion see Zimmerman 2004). Nothing in this paper turns on the debate over whether intrinsic value or final value should be the focus of axiology. For the sake of familiarity and historical continuity, I continue to use the term 'intrinsic value'. I also assume, though nothing depends on it, that states of affairs rather than concrete objects are the fundamental bearers of value.

[2] Fitting-attitude analyses are, however, compatible with a naturalistic reduction of the deontic notion they contain, a notion that, for the purposes of this paper, will be taken as primitive.

cannot, then, define the ultimately good or desirable otherwise than by saying that it is that of which we should desire the existence if our desires were in harmony with reason . . . ."[3] Many other philosophers have since defended, or at least seriously considered, more or less the same idea.[4] Here is how C. D. Broad (1930: 238) puts it: "I am not sure that 'X is good' could not 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." Though he doesn't mention Broad, T. M. Scanlon (1998: 97) has recently endorsed a view very much like Broad's. Scanlon's so-called buck-passing account of value has been receiving a lot of attention.

It is fitting to want such an analysis to be true. Fitting-attitude analyses do at least the following valuable things:

i. They reduce. I believe in value, and I believe in reasons to have certain attitudes. This view reduces the one to the other. So instead of having to believe in both of these things as fundamental, irreducible features of the world, I have to believe, fundamentally, in at most only one—reasons—and I get the other for free.

ii. They demystify. Some people are skeptical of intrinsic value, but fewer people are skeptical of the notion that we ought to have certain attitudes. We all, for instance, believe that there are certain things we ought to believe. So we are familiar with the idea that certain attitudes are required. If we see that facts about intrinsic value are nothing more than facts about what attitudes there are reasons to have, we may no longer find intrinsic value so mysterious.

iii. They explain why it is confused to wonder whether there is any reason to promote the good. If something is intrinsically good, then, given

[3] So Moore was evidently mistaken in taking Sidgwick to have agreed with him about the indefinability of goodness; Moore wrote, " 'Good,' then, is indefinable; and yet, so far as I know, there is only one ethical writer, Prof. Henry Sidgwick, who has clearly recognised and stated this fact" (1903a: § 14). Perhaps Moore was misled by the fact that Sidgwick held that "The term 'ought,' as used in moral judgments . . . is unanalysable" (1884: p. xvi). See also Sidgwick 1884: 27 or 1907: 25.

[4] See Brentano 1889: 18, John Maynard Keynes 1905, C. D. Broad 1930: 238, Richard Brandt 1946: 113 and 1959: 159, A. C. Ewing 1947: 152, J. O. Urmson 1968: 58–9, John McDowell 1985: 118, Roderick Chisholm 1986: 52, David Wiggins 1987 : 189, Allan Gibbard 1992: 980, Elizabeth Anderson 1993: 1–2, Richard Kraut 1994: 45, Noah Lemos 1994: 12, T. M. Scanlon 1998: 95, 97, Thomas Carson 2000: 158–9, Michael Zimmerman 2001: ch. 4, Derek Parfit 2001: 20, Wlodek Rabinowicz and Toni Rønnow-Rasmussen 2004, Jonas Olson 2004, Jussi Suikkanen 2004, and Philip Stratton-Lake and Brad Hooker 2005. Rabinowicz and Rønnow-Rasmussen 2004: 394–400, from which some of the above citations are drawn, contains a helpful historical summary. See also Carson 2000: 160–2, from which some other of the above citations are drawn. Keynes 1905 is unpublished; for discussion, see Baldwin 2006.

a fitting-attitude analysis, it follows automatically that there is reason to want it to occur.

1.2. A Formal Problem: The Wrong Kind of Reasons Problem

Fitting-attitude analyses face a certain sort of formal objection.[5] I call this objection 'formal' because it seems to me it doesn't strike at the heart of the basic idea of fitting-attitude theories; it instead guides us in how best to formulate the view, or in how best to understand the view we have already formulated.

One sort of problem case involves *desiring a bad thing for the bad thing's good effects.* According to fitting-attitude theories, to say that something is intrinsically good is to say that there is reason to want it to occur. But suppose I have a cut that needs to be cleaned with alcohol. Suppose that only if I am feeling the sting of alcohol in the cut is it being cleaned. I therefore want to be feeling this painful sensation. Given that avoiding infection requires feeling this sensation, it is sensible to want to be feeling it. But then the fitting-attitude theory implies the absurd claim that this pain is *intrinsically* good.

This case teaches us that fitting-attitude analyzers have *intrinsic desire*, or desire for something for its own sake, in mind when they say that to be intrinsically good is to be fit to be desired. A fitting-attitude theory restricted to intrinsic desire doesn't imply that my pain is intrinsically good (since I don't have reason to want it for its own sake, just for what will accompany it).

Another sort of case involves *desiring a bad thing for the desire's good effects*. Suppose a demon offers to end world poverty if only you will intrinsically desire something bad, like that Tiger Woods gets a terrible headache later today. Surely you ought to get yourself to desire for its own sake (assuming that's even possible) that Tiger Woods gets the headache later today. But still, Tiger's having the headache is not intrinsically good.

This case teaches us that that fitting-attitude analyzers have *object-given reasons*, rather than *attitude-given reasons*, in mind when they say that something is good just in case there is reason to desire it intrinsically.[6] As

[5] See D'Arms and Jacobson 2000, Rabinowicz and Rønnow-Rasmussen 2004, which introduces the expression 'Wrong Kind of Reasons problem', Jonas Olson 2004, Stratton-Lake 2005, Rabinowicz and Rønnow-Rasmussen 2006, and Danielsson and Olson 2007. Danielsson and Olson (2007) point out that Moore (1903b) may have put forth a version the Wrong Kind of Reasons problem as well.

[6] This distinction is drawn in Parfit 2001: 21–2, though not with these exact labels.

the case illustrates, sometimes reasons to have an attitude derive not from the *object* of the attitude, or from what that object would bring about, but from the *attitude itself*, or from what having it would bring about. In more typical cases, when we have reasons to have some attitude towards some object, the reasons are provided by the nature or effects of the object, as when, for example, we all have reasons to want world poverty to end. These reasons derive not (or not wholly) from any effects of having the attitude but instead from the badness of poverty, or from the non-evaluative features of poverty that make it bad.[7]

1.3. Overview of My Argument

My argument against fitting-attitude analyses of value is indirect. The first part of it says that anyone endorsing a fitting-attitude analysis of intrinsic value ought also, for the sake of theoretical unity, to endorse a fitting-attitude analysis of a closely related but distinct concept: the concept of intrinsic value *for a person*, that is, the concept of *welfare*. It would be surprising to learn that whereas intrinsic value *simpliciter* is analyzable in terms of reasons to have an attitude, intrinsic value for a person is not at all analyzable in this way, and is perhaps instead unanalyzable. This suggests (though admittedly does not entail) that fitting-attitude analyses of intrinsic value *simpliciter* were wrong all along.

The second part of my objection consists in arguing directly against fitting-attitude analyses of welfare. This is what the rest of the paper is about. Even if one rejects the first part of the argument, one can understand this paper to be arguing just against fitting-attitude analyses of welfare. This is not mere shadow boxing, for fitting-attitude analyses of welfare have been defended independently.[8]

[7] Rabinowicz and Rønnow-Rasmussen (2004: 404–8) maintain that restricting a fitting-attitude analysis to object-given reasons still doesn't avoid the Wrong Kind of Reasons problem. They complain that it is difficult to spell out the conditions under which a reason is object- rather than attitude-given (and that certain proposals for spelling it out don't help the fitting-attitude theorist). That may be correct, and I agree it would be best for a fitting-attitude theorist to be able to spell out the conditions. But since we have an antecedent, intuitive grasp of when reasons are object- rather than attitude-given, a grasp that does not depend upon intuitions about value (the analysandum of fitting-attitudes analyses), I think the appeal to the object-given/attitude-given distinction solves the problem well enough. Even if I'm wrong about this, however, granting that the problem is solved only helps the theory this paper is attacking.

[8] e.g. by Sidgwick (1907), Stephen Darwall (2002), and Rønnow-Rasmussen (forthcoming). Their theories are quoted below and briefly discussed.

2. FITTING-ATTITUDE ANALYSES OF WELFARE

2.1. Value *Simpliciter* vs. Welfare Value, or Impersonal vs. Personal Value

On the one hand there is welfare, or well-being, or personal value, or prudential value, or value for a person (or some other subject), or what makes for benefit and harm, or what makes a life in itself worth living. On the other hand, there is value *simpliciter*, or impersonal value, or "value for the world," or what makes the world a better or worse world. Although welfare is a kind of intrinsic value, it is also a *relational* kind of value. We express the idea when we say that something would be *good for* someone. Welfare value is intrinsic because we are saying the thing is good in itself for the person, and not merely good for what it leads to for the person. It is relational because it is a relation between the thing that's good and the person for whom it's good.

Both impersonal and personal value are *non-moral* kinds of value. When we say that something is good in one or the other of these ways, we do not mean the thing is morally good (as only an agent or an action can be). Personal value is rather the kind of value that makes a person's *life go better*. And impersonal value is the kind of value that makes an *outcome* better, or the *world* a better place, or, as we might say, makes *things go better*.

Despite their similarities, these concepts are independent from one another. We can believe that something is good for someone (that its presence makes his life go better) without believing that it is good (that its presence makes the world better, or makes things go better). Consider, for instance, the prospering of the wicked. We can believe Ted Bundy's enjoying something makes his life go better for him without thinking that this enjoyment makes things go better. If we believe injustice to be impersonally intrinsically bad, we might even think it makes things go worse when someone like Ted Bundy is having his life go better.

Likewise, we can think something helps make the world a good world even though it doesn't help make anyone's life good. If we think justice is impersonally intrinsically good, we may think that Bundy's getting the suffering he deserves is intrinsically good, even though this isn't intrinsically good for anyone. We may also think that equality, beauty, virtue, excellence, or noble action is intrinsically good without being intrinsically good for anyone.

The point of all this for our larger purpose here is that even if we have an acceptable analysis of value, we still have to deal with the concept of welfare. We have to be sure that our analysis of value "carries over" to our analysis of welfare, since it seems implausible, and would be theoretically unsatisfying, to be forced to say that these two kinds of value have radically different natures. But, as I will argue momentarily, fitting-attitude analyses do not in fact carry over to welfare.

2.2. Fitting-Attitude Analyses of Welfare

Sidgwick (1907: 112) held that a fitting-attitude analysis applies to welfare value as well as to value *simpliciter*. He interprets " 'ultimate good on the whole for me' to mean what I should practically desire if my desires were in harmony with reason, assuming my own existence alone to be considered" (1907: 112). More recent attempts to carry over fitting-attitude analyses to welfare are made by Stephen Darwall (2002) and Toni Rønnow-Rasmussen (forthcoming).[9] Darwall writes: "what it is for something to be good for someone *just is* for it to be something one should desire for him for his sake, that is, insofar as one cares for him" (2002: 8–9). Rønnow-Rasmussen writes: "An object has personal value for a person *a* if and only if there is reason to favor it for *a*'s sake (where 'favor' is a place-holder for different pro-responses that are called for by the value bearer)."[10] Since one obvious way to understand the expression 'for *a*'s sake' is as meaning the same as 'for *a*'s benefit' or 'for *a*'s welfare', the latter two theories appear circular. But both Darwall (pp. 1–2) and Rønnow-Rasmussen (pp. 14–17) suggest that desiring or favoring something for someone's sake is instead a *way of desiring*, and is therefore (if I understand them correctly) a purely psychological notion, not one whose analysis requires appeal to the notion of welfare.

Before presenting my argument against views such as these, I want to emphasize what seems to me to be an advantage of Sidgwick's version of the theory. It is the fact that, on Sidgwick's view, for something to be good for someone is for it to give *that person*, and not necessarily anyone else, the reasons to have the desire. Cases of the prospering of the wicked reveal the benefit this feature brings. As noted above, it might be a bad thing (or at least fail to be a good thing) for the wicked to prosper, even

[9] Page references for Rønnow-Rasmussen's paper relate to the manuscript available at <http://www.fil.lu.se/publicationfiles/pp88.pdf>.

[10] I should point out that Rønnow-Rasmussen does not mean by 'personal value' what I mean by it. As I use the term, personal value is the same thing as welfare, but as Rønnow-Rasmussen uses it, welfare is just one of two kinds of personal value (Rønnow-Rasmussen, forthcoming: 16). But since welfare is a species of personal value on his view, Rønnow-Rasmussen's theory commits him to the thesis my argument attacks: that if something has welfare-value for *a*, then there is reason to favor it for *a*'s sake.

though their prospering is good *for them*. I think that Ted Bundy's enjoying something is good for him, but it is reasonable to think that this enjoyment fails to be good *simpliciter*.[11] It is plausible that the fact that he would enjoy something doesn't give the rest of us any reason to bring it about—not even an outweighed reason. But Bundy himself still has reason to bring it about—he, after all, is the one who would benefit. So whereas Rønnow-Rasmussen's version of the view implies that if something would be good for Bundy, then *everyone* has reason to favor it, Sidgwick's version more plausibly implies only that Bundy himself has reason to favor it.[12]

Here, then, is the target theory. I include the necessary part analyzing negative welfare value since my argument is more naturally presented as against this part of the theory. I also intend this theory to reflect the solutions discussed earlier to the formal problems for fitting-attitude analyses generally.

FA1 x is intrinsically good for S iff x itself gives S reason to intrinsically desire x for S's sake;

x is intrinsically bad for S iff x itself gives S reason to be intrinsically averse to x for S's sake.

The expression 'x itself gives S reason' is meant to indicate that the relevant reasons are object- rather than attitude-given. Though I formulate the view in terms of reasons (as many fitting-attitude analyzers nowadays do), I will, in what follows, for stylistic variation, use other related notions, like that of an attitude being *fitting*, or *appropriate*, or *warranted*, or *rational*, or one that *makes sense*, or one we *ought* to have. I also often leave the 'intrinsically' qualifier for the attitude implicit. I formulate these theses using desire and aversion since fitting-attitude analyzers commonly use these notions, but nothing hangs on this (and, in what follows, I will, for stylistic variation, make free use of other pro- and con-attitudes).

The expression 'for S's sake' is important. Without it, FA1 might imply that anything impersonally good (such as, e.g., Tiger Woods's deservedly enjoying a rousing ovation) is also good *for every subject*, since perhaps we all have reason to want impersonally good things to occur. But since we don't want, and don't have reason to want, these things "for our own sakes," FA1 avoids this implication.

Notice that FA1 involves two "sakes." If something would be good for someone, then she has reason to want it intrinsically for her own sake. But to want something intrinsically is to want it for its own sake. Thus she has reason to want it for its own sake for her own sake. She has reason to want it for its own sake—rather than for the sake it what it would lead to—and also has reason to want it for her own sake—rather than for my sake or for no one's sake at all.

3. WHY FITTING-ATTITUDE ANALYSES OF WELFARE FAIL

My argument against FA1 has to do with time. It is based on the idea that while an event's value for a person is unchanging, the attitudes he has reason to have towards such an event can change over time. We can illustrate this idea using Derek Parfit's ingenious case *My Past and Future Operations* (1984: 165):

> I am in some hospital, to have some kind of surgery. Since this is completely safe, and always successful, I have no fears about the effects. The surgery may be brief, or it may instead take a long time. Because I have to co-operate with the surgeon, I cannot have anaesthetics. I have had this surgery once before, and I can remember how painful it is. Under a new policy, because the operation is so painful, patients are now afterwards made to forget it. Some drug removes their memories of the last few hours.
>
> I have just woken up. I cannot remember going to sleep. I ask my nurse if it has been decided when my operation is to be, and how long it must take. She says that she knows the facts about both me and another patient, but that she cannot remember which facts apply to whom. She can tell me only that the following is true. I may be the patient who had his operation yesterday. In that case, my operation was the longest ever performed, lasting ten hours. I may instead be the patient who is to have a short operation later today. It is either true that I did suffer for ten hours, or true that I shall suffer for one hour.

[11] This point applies even if some objective list theory of welfare is correct (I appeal to enjoyment here and throughout only because it is the least controversial example of a human good). Supposing, for instance, that *experiencing true beauty* is one of the great human intrinsic goods, then, intuitively, it will be good for Bundy to experience true beauty even if it fails to be a good thing that Bundy gets to do this. It is worth pointing out, however, that there is a problem here regarding the idea that there is a close connection between welfare and virtue. If only the virtuous can have good things happen to them, then it may be that anytime someone has something good happen to him, this is also a good thing (since it will automatically be a case of a good person getting something good for him). I thank Mark LeBar for discussion here.

[12] Rønnow-Rasmussen is aware of this objection, and discusses some strategies available to proponents of his view (Rønnow-Rasmussen, forthcoming: 24–5). But I think the Sidgwickian approach presented above is better than the ones Rønnow-Rasmussen considers. Darwall's care condition might seem to enable his theory to avoid this objection. But I think including this condition is in other ways problematic: it threatens to make the analysis circular, as I argue briefly in Heathwood 2003; and, as Rønnow-Rasmussen (forthcoming, 23) points out, it appears to commit the analysis to a subjective, attitude-based theory of practical reason. In any event, the argument I present below still applies to both Rønnow-Rasmussen's and Darwall's original formulations, as I indicate in the next footnote.

I ask the nurse to find out which is true. While she is away, it is clear to me which I prefer to be true. If I learn that the first is true, I shall be greatly relieved.

I make two additional suppositions. First, assume that, in the case, Parfit has his preference—his "bias towards the future"—in virtue of two facts: that he is strongly averse to, or strongly disfavors, his suffering for one hour tomorrow (as the nurse says he might); and that he has no aversion at all to the idea of his having suffered for ten hours yesterday. Thus, I am assuming (but only for now and only for simplicity) that Parfit's bias towards the future is *extreme*: he cares not at all about his own past suffering; he would not "buy" any reduction in past suffering, however large, in exchange for any increase in future suffering, however small. Second, suppose that, as a matter of fact, Parfit is the patient who had his operation yesterday, the operation that lasted ten hours.

Here is the argument against FA1:

(1) Parfit's suffering for ten hours yesterday was intrinsically bad for Parfit.

This is undeniable. We don't need to be hedonists to think suffering is bad for those who suffer—this is rather a datum that any theory of welfare must respect. Note that premise 1 is not saying that Parfit's suffering for ten hours yesterday is *all things considered* or *on balance* bad for Parfit. We can assume that this suffering is all things considered good, due to the good effects of the surgery, which outweigh the badness of the suffering and which would not have occurred had Parfit not suffered.

(2) If Parfit's suffering for ten hours yesterday was intrinsically bad for Parfit, then if FA1 is true, then Parfit's suffering for ten hours yesterday gives him reason to be intrinsically averse to that ordeal for his own sake.

This premise just applies FA1 to the case at hand. If in fact the event of Parfit's suffering for those ten hours really has negative value for him, then, by the second clause of FA1, we can conclude that it gives him reason to be averse to the fact that it happened.[13]

(3) But it is false that Parfit's suffering for ten hours yesterday gives him reason to be intrinsically averse to that ordeal for his own sake.

This is a crucial premise. But it should be intuitively compelling once one appreciates what it is saying. Parfit is being completely reasonable in preferring that his pain be in the past. In fact, even his no longer caring

[13] Rønnow-Rasmussen's analysis likewise implies that Parfit's suffering gives him reason to be averse to it, since, according to this analysis, Parfit's suffering gives everyone reason to be averse to it. Darwall's analysis also implies this, so long as we stipulate that Parfit cares for himself.

at all that it occurred is perfectly fitting—not at all inappropriate. Why should he care about it now? No reason—it's over and done with. When things become past, the reasons they provide can change.

But this isn't true for value; whether an event is intrinsically good or bad for a person doesn't change. It will always remain true that Parfit's ordeal was bad for him (just as it will always remain true that Parfit's ordeal actually occurred). His life is made worse as a result, and there's nothing anyone can do now to change that.

From these claims it follows that

(4) Therefore, FA1 is not true.

Since fitting-attitude analyses of welfare fail, the fitting-attitude approach to value generally looks less attractive. It is more reasonable to suppose that a unified account—one according to which the analysis of impersonal value, if any, carries over to the analysis of welfare—is correct than that personal value and impersonal value have radically different natures.

4. OBJECTIONS AND REPLIES

4.1. The Unfittingness of the Extreme Bias towards the Future

Having no bias at all towards the future does seem crazy, but the extreme bias you stipulate for Parfit is too extreme. Even though it's over and done with, it's still appropriate to be at least a little bit against, for your own sake, the fact that you underwent some terrible ordeal (even when you have no memory of it). After all, it was a terrible ordeal, and you really underwent it.

Suppose Parfit was deciding not between ten hours in the past and one hour in the future of equally intense suffering but between ten years of the most horrific torture in the past and one second of a barely noticeable pain later today. Maybe reason demands that he prefer the latter.

I happen not to be convinced that reason demands that he prefer the latter. So long as we're careful to keep in mind that the ten years of past agony has no bad side-effects—there are no memories of it (which would be bad to have), no post-traumatic stress, no injuries, no concomitant loss of goods[14]—then if Parfit insists, "I don't see why I should care at all about this ordeal that I evidently underwent but that is now over and done with," I don't think we could convince him that he ought to care about it. And I don't think his stubbornness would be unreasonable.

[14] To imagine this properly, we can imagine that, had Parfit not been tortured, he would have been in a coma for those ten years.

But this question is irrelevant anyway. This is because the extreme bias was stipulated only for the sake of simplicity of initial presentation. An analogous argument goes through without assuming that the extreme bias is appropriate.

To see this, first notice that FA1 is an oversimplification. Goodness and badness come in degrees—one good thing can be better than another good thing—but FA1 doesn't reflect this. A complete analysis of personal value would specify how *degree* of goodness or badness depends upon degree of fittingness of attitude, or upon fittingness of degree of attitude. The complete theory would analyze not the notion corresponding to 'x is intrinsically good for S' but the one corresponding to 'x is intrinsically good for S *to degree n*'. One natural view makes use of the notion of strength of reason, as follows:

FA2 x is intrinsically good for S *to degree n* iff x itself gives S reason *of strength n* to intrinsically desire x for S's sake;

x is intrinsically bad for S *to degree n* iff x itself gives S reason *of strength n* to be intrinsically averse to x for S's sake.

Since the better or worse something is, the more reason there is to be for or against it, fitting-attitude analyzers of welfare are committed to something relevantly like FA2.[15] But, since one hour of suffering is intrinsically better (in other words, less intrinsically bad) for the sufferer than ten hours of equally intense suffering, FA2 implies that Parfit's past pain gives him more reason to be averse to it than the future pain would give him to be averse to *it*.

But that's not right. The past pain is not cause for greater alarm than the future pain would be. Parfit's future pain, despite being less bad, provides more reason to be against it.

In response to this, a defender of FA2 might insist

Parfit's suffering for ten hours yesterday does *give him more reason to be averse to that suffering than the future suffering would give him. Just considering the pains themselves, Parfit has more reason to be averse to the past pain (it's a greater pain, after all). But, crucially, this does not commit me to the claim that Parfit should, all things considered, prefer that his operation be tomorrow. For taking into account all reasons, Parfit has more reason to be averse to the future pain, since there are (as this very case illustrates)* time-related reasons, *and these time-related reasons tilt the balance so far in the other direction that, all things considered, Parfit has more reason to be averse to the future pain.*[16]

I continue to think that it is just false that the past ordeal gives Parfit more reason to be averse to it than the future ordeal does. The past ordeal is over and done with, so it no longer merits much concern (even if, as we are now granting, it does merit a little). And it certainly merits less concern than the future ordeal would merit if it were looming.

But I don't need to rest my response upon this claim, because the appeal to these time-related reasons brings with it new problems for the fitting-attitude theorist. In making this appeal, the defender of FA2 is claiming that although Parfit's past pain *provides* more reason than would the future pain, Parfit nonetheless would *have* more reason to be averse to the future ordeal, due to these time-related reasons. But if Parfit would have this additional reason, it would have to come from somewhere. Something would have to provide this additional reason, something along the lines of the following fact (where t3 is the time of the future operation):

F *that Parfit suffers to degree 10 at t3, and t3 is in the future.*

On this picture, the reasons a pain provides are always proportionate in strength to the amount of pain in the pain, no matter the pain's temporal location. And then the further fact (like fact F) that some pain is in the future provides further reason.

Perhaps this picture is correct, but it cannot be combined with a fitting analysis like FA2. For on any fitting analysis of welfare, not only is it true that whenever there is welfare, there are reasons of a certain sort, it is also true that whenever there are reasons of that sort, there is welfare. If some fact such as F gives someone a reason of some strength to be intrinsically averse to it for his own sake, then, given FA2, it follows that that fact is intrinsically bad for the person to that degree, and makes his life that much worse.

But the idea that F would add disvalue to Parfit's life over and above the disvalue contributed by the suffering is implausible. To see this, consider a bad life that is at its midpoint and whose future is a perfect duplicate its past. This should allow us to say to its subject, "Although your future, like your past, won't be any good, at least it won't be worse than your past was." But, on the present suggestion, this won't be true. We would instead be required to say, "although your future is indiscernible from your past with respect to its non-evaluative features, it will be far worse

[15] So, for example, Darwall would be committed, if we were to mirror his formulation, to the view that what it is for something to be good for someone *to some degree* just is for it to be something one *to that degree* should desire for him for his sake, that is, insofar as one cares for him. This theory makes use of the idea of stringency of (prima facie) obligation.

[16] I am grateful to Ben Bradley, Jens Johansson, and Doug Portmore for (independently) raising objections along these lines.

than your past was." Defenders of FA2 who adopt the above appeal to time-related reasons will be required to say this because, as we all agree, the subject of this life would have far more reason to be averse to his future than to his past. Given FA2, this will imply that the future is far worse. But clearly it isn't—this person's past is just as bad his future.

4.2. The Reasons Everyone Has to be Averse to Anyone's Suffering

Parfit's past suffering isn't just bad for him, it's a bad thing. Thus we all have reasons to wish it didn't happen. If one were to feel bad about what Parfit went through, that would be fitting (just as it is fitting for anyone now to be disturbed, say, over what one of Ted Bundy's victims went through). But surely Parfit can take up the same "impersonal point of view" towards his past self, and feel bad that there was this person who underwent this terrible ordeal. Such an attitude would be fitting. So premise 3 is false. Parfit has the same reasons we all have to be averse to his suffering yesterday.

I agree that, since Parfit's suffering was impersonally bad, he has the same reasons we all have to be averse to it. But this doesn't contradict premise 3. To contradict premise 3 in the way intended, it needs to be shown that Parfit has reason to be averse to his suffering yesterday *for his own sake*. But, for his own sake, why should he care at all about his past ordeal? It's over and done with.[17]

This reply can be made clearer if we pretend that Parfit is extremely wicked, so that any suffering he undergoes gives no one any reason to be averse to it. Given this supposition, his past suffering is no longer a bad thing (although it is still, of course, bad for him). Since "from the impersonal point of view" no one has reason to feel bad about Parfit's ordeal, he doesn't either, from the impersonal point of view. Thus the only reasons the past suffering could provide anyone to be intrinsically averse to it would be the reasons it provides Parfit himself to be against it "from the first person point of view." But, even if it provides him some such reason to be against it, it doesn't, as FA2 implies, provide more reason to be against it than the future ordeal would provide him.

[17] Here I am again assuming the fittingness of the extreme bias. But this is just for simplicity. Speaking just about what Parfit should want for his own sake, he has stronger reason to desire that he be the patient whose ordeal is over than that he be the patient whose ordeal is looming (even if, since an extreme bias is irrational, he has some reason to be averse for his own sake to both). In what follows, I will continue to assume, for simplicity, the fittingness of the extreme bias.

4.3. The Fittingness of Having No Temporal Bias

True, it is fitting to have a bias towards the future, but it is also fitting to have no temporal bias at all. Each is rationally permissible. Therefore, it would be ok for Parfit to prefer to be the patient whose operation is later today. But the arguments against FA1 and FA2 assume that having no such temporal bias is irrational.

I happen not to be convinced that reason permits Parfit to prefer to be the patient whose operation is later today. Were we in Parfit's shoes, then while the nurse is away, we all would reasonably anticipate, with dread, the possibility of being the patient whose operation looms. If Parfit had no temporal bias, then he would look backward, with a backward-looking analog to dread, to the possibility of being the patient whose operation is over. This doesn't just seem odd, it seems like a mistake. We'd say, "Look—don't you get it? If you're the patient whose operation was yesterday, then your suffering is *over and done with. It's a thing of the past*. Stop getting worked up about it. That doesn't make any sense."

But this question is irrelevant anyway. My point stands even if we grant the permissibility of having no temporal bias. For defenders of a fitting-attitude analysis of welfare like FA2 are committed to the claim that Parfit's past ordeal, since it is worse, gives him more reason to be averse to it than would be given by the lesser, future ordeal. So even if the defender of FA2 somehow nevertheless allows that it is rationally permissible for Parfit to prefer in the temporally biased way he does, she must claim that Parfit fails to prefer in the way that he has *most reason* to prefer. FA2 entails that the balance of reason tilts in favor—very strongly in favor, in fact—of Parfit preferring that his ordeal be in the future (since the possible future ordeal, according to FA2, provides significantly less reason to be averse to it than does the possible past ordeal, and since no other reasons are operating). So if Parfit's attitudes are to be in *full harmony with reason*, he needs to be much more strongly opposed to the thought of being the patient whose operation is over and done with.

But that is not true. And the implausibility of this claim is in no way compensated for by the concession that Parfit's bias towards the future, although way out of whack with what he has most reason to prefer, is nevertheless *permitted* by reason.

It would be nice if we had an argument (in addition to the appeals to intuition) for the claim that having a bias towards the future is fully rational. But I don't think this thesis about rationality can be explained in terms of, or subsumed under, any more general claim about rationality. It seems to me we've reached a brute fact about rationality. One can try to

explain why the way Parfit prefers, and the way we all prefer, is perfectly rational by pointing out that Parfit *still has to undergo* the pain if his operation is later today, but the pain *is over and done with* if his operation was yesterday. But this adds nothing. It just repeats in different words what needed explaining.[18] Why is it preferable for a pain to be over and done with? I'm afraid the only answer may be: it just is.

Though the rationality of the bias towards the future may be inexplicable, it is important to note the implausibility of a tempting sort of debunking explanation of our intuitions in favor of its rationality. The debunker might claim that we all intuit that the bias towards the future is rational only because we all have the bias; our having the intuition is thus better explained by its being self-serving than by its being true. This argument is unpersuasive because there are other biases that are ubiquitous but that we nevertheless intuit to be positively *irrational*, such as the so-called bias towards the near—our tendency to care more about our nearer future than about our further future, as when we prefer to delay suffering and hasten enjoyment.

4.4. Time, Tense, and Temporal Indexing Strategies

Something funny is going on in your argument with tense. FA1 and FA2 seem to be stated in the present tense, but the welfare attributions in your argument are stated in the past tense. As a result, it is not clear that

(2) *If Parfit's suffering for ten hours yesterday* was *intrinsically bad for Parfit, then if FA1 is true, then Parfit's suffering for ten hours yesterday* gives *him reason to be averse to that ordeal for his own sake.*

is true. Since FA1 and FA2 are theories about what it means to say something is *bad for someone, they imply nothing concerning claims to the effect that something* was *bad for someone.*

Fitting-attitude analyses, I was assuming, were meant *tenselessly*. Surely fitting-attitude analyzers mean the theory to be general, so that it applies to all value and welfare judgments, irrespective of their tense.

Still, the theories do seem to be incomplete, since people have reasons to have attitudes at times. Just as FA1 was incomplete in failing to include degree indices, both FA1 and FA2 are incomplete in failing to including temporal indices. Perhaps the completed fitting-attitude analysis of welfare is

FA3 x is intrinsically good for S *at time t* to degree n iff x itself gives S reason *at t* of strength n to intrinsically desire x *at t* for S's sake;

[18] Cf. Parfit 1984:178.

x is intrinsically bad for S *at time t* to degree n iff x itself gives S reason *at t* of strength n to be intrinsically averse to x *at t* for S's sake.[19]

FA3 is not *ad hoc*. The motivation for including the temporal indices is independent of the debate at hand. Perhaps this natural way to make the theory complete will also help it avoid my argument. This would show that my argument has force only against an oversimplification of a fitting-attitude analysis of welfare.

FA3 does not imply, as FA1 appears to, that Parfit now has reason to be averse to his past ordeal. The time at which Parfit's suffering for those ten hours yesterday has disvalue for him is *during those ten hours*. FA3 therefore implies only that Parfit *had* reason to be averse to his past ordeal, during those ten hours, which surely he did.[20] For similar reasons, FA3 does not imply, as FA2 appears to, that Parfit has more reason to be averse to his past possible long ordeal than his future possible shorter ordeal. This is because FA3 doesn't imply that Parfit has any reason to have any attitude about either possible ordeal.

But FA3 faces new problems. Suppose that Parfit will in fact undergo the future operation. We all agree that this future event gives Parfit reason now to be against it (intrinsically and for his own sake). It is reasonable for Parfit now to be dreading the fact that he will undergo it. But on FA3, whenever there are reasons at a time to have certain desires at that time, there is value at that time. So FA3 implies, absurdly, that Parfit's future suffering is bad for him *now*. But it's not bad for him now—it will be bad for him when it is occurring.

Let me be clear about what FA3 implies here. It implies that Parfit's future suffering is *intrinsically* bad for him now, and that is what's implausible. It is at least conceivable that Parfit's future suffering is *extrinsically* bad for him now, due to the anxiety it might be thought to give rise to now. I suppose this would be true if the following "backtracking" counterfactual were true: if Parfit weren't going to be suffering later today, then he wouldn't be feeling anxious right now. But it is not possible that Parfit's future suffering is *intrinsically* bad for him now. This is because the view that future suffering is intrinsically bad now (or that any future evils are intrinsically bad now) implies that one's present days are *made worse* by the existence of these future evils. It would thus imply, absurdly, that when someone asks, "How

[19] Views like this have been suggested to me by Campbell Brown and by Stephan Torre (personal correspondence).

[20] It may be more accurate to say that *at each individual moment* of Parfit's suffering, the suffering he experienced at that moment was bad for him at that moment; and, likewise, that at each individual moment of the ordeal, Parfit had, at that moment, reason to be averse to the suffering he was experiencing at that moment.

was your day?", to answer accurately you need to consider not only what happened to you that day, but everything that will ever happen to you in all of your remaining days.

FA3 has another defect. It suggests that for any state of affairs that is good for a person, there is some particular time at which it is good for him. But some philosophers have independently endorsed certain normative theses about welfare that are hard to reconcile with this. For example, David Velleman (1991) claims that the narrative structure of a life can impact how good it is for the person (others before Velleman, including Brentano himself, have made similar claims[21]). So the state of affairs of your life having such-and-such structure could be good for you—it could make your life better than it would have been. But there doesn't seem to be any particular time at which the state of affairs of your life having this structure is good for you.

Another instance is Thomas Nagel, who discusses examples meant to illustrate that "while [a] subject can be exactly located in a sequence of places and times, the same is not necessarily true of the goods and ills that befall him" (1970: 77). I don't know whether Velleman's and Nagel's normative views are correct, but it would be better if our metaethical theory didn't rule them out right off the bat, making them conceptually confused, or false by definition.

An advocate of FA3 might reply to the Velleman case that for it to be true that there is no *particular* time at which the state of affairs of your life having its nice structure is good for you, it is enough that it be good for you *at every time*.[22] FA3 would then imply that you have reason, at every time, to be in favor of your life's structure. Though this claim about reasons is plausible, the associated claim that your life's structure is good for you at every time is not. It implies that each moment of your life is made better by the structure had by your whole life. It would thus imply, absurdly, that when someone asks, "How was your day?" to answer accurately you need to consider not only what happened to you that day, but also the overall structure of your whole life.

Perhaps, though, now that we have seen the problems with one temporal indexing strategy, we can use what we have learned to rig up an analysis that will spit out the results we want, results that harmonize with the idea that present and future goods and evils give us reasons (or

[21] e.g. Slote (1983) and Chisholm (1986). See also Lemos 1994 and Carson 2000. Carson 2000 contains a useful overview of the views of some of these philosophers on this topic.

[22] Campbell Brown and an anonymous referee have both suggested this reply. Perhaps a similar reply might be made to the Nagel cases as well.

reasons to a degree) that past goods and evils do not. Here is one such proposal:

FA4 x is intrinsically good for S to degree n iff x itself gives S reason of strength n to intrinsically desire x *before or during the time at which x occurs* for S's sake;

x is intrinsically bad for S to degree n iff x itself gives S reason of strength n to be intrinsically averse to x *before or during the time at which x occurs* for S's sake.[23]

According to FA4, if some event that either is occurring or will occur is bad for someone, then this event gives him reason to be intrinsically averse to it for his own sake. But if some valuable event for someone is over and done with, then FA4 does not imply that it now gives him any reasons to have any attitude. FA4 does imply that he *did* have such reasons in the past, before and during the time of the event. But FA4 is compatible with the idea that Parfit now has no reason at all to care about his past suffering.

But FA4 faces new problems. For one thing, it is *ad hoc*. It includes complicated epicycles in the form of disjunctive temporal qualifiers only to get the right result to a specific objection. Since *ad hoc* theories are less likely to be true, we should be dubious of FA4.

But more importantly, FA4 is not even extensionally adequate. To see why, first note an interesting distinction between certain kinds of alleged goods and evils, one that relates importantly to the bias towards the future. When it comes to our own pleasure and pain, the bias towards the future is ubiquitous and sensible. But there are other putative goods and evils about which we are not temporally biased. One example is behind-the-back ridicule. Some philosophers have argued that when you are ridiculed behind your back, this is bad for you, independently of whether you ever find out about it.[24] Though we never know about it, our lives are made

[23] Theories like this have been suggested to me by Michael Huemer and by Elizabeth Harman. It is interesting to compare FA4 to the following theory Sidgwick discusses (1907: 111–12): "a man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time." But there is no reason to think Sidgwick states this theory of welfare in terms of "a man's *future* good" in an attempt to avoid our time-related worries. I suspect he just finds it natural to state it this way, since it brings to mind the perspective of the deliberating agent.

[24] Nagel (1970: 76) may be the most prominent advocate of this view. Other examples of non-experienced evils according to Nagel include betrayal, deception, being despised, and having one's will ignored after one's death (1970: 76). See also Kagan (1998: 34–5). It is worth noting that the idea that ridicule, betrayal, hatred, and deception are bad independently of our awareness of them isn't just an intuition. It provides a simple and satisfying explanation for why their discovery is upsetting. Likewise, the idea

worse when such things happen to us. Maybe that's true, but, interestingly, when it comes to such evils, we do not prefer that their instances be in our past. Suppose we learn that it is either true that we did suffer ten ridiculings last week, or true that we shall suffer one ridiculing this week. If later we learn that the first is true, we shall *not* be greatly relieved. We simply (and reasonably) prefer fewer ridiculings, no matter their temporal location. Though this is not so with the "experienced evils" we have been discussing up until now, we have no bias towards the future when it comes to non-experienced evils.[25]

Non-experienced evils make trouble for FA4 in at least two ways. First, consider the possibility *pre-vital harm*. The notion of *posthumous harm* has been widely discussed and defended, but could there ever be pre-vital harm? Could there be an event that occurs before a person begins to exist that is nevertheless intrinsically bad for that person?[26] If there are non-experienced evils such as behind-the-back ridicule, I don't see how one can rule it out. If it is bad to be ridiculed independently of whether one could ever find out about it, then it should still be bad even if it occurs after its victim is dead.[27] And, likewise, it should still be bad even if it occurs before its victim is born.

So, it is a live option in normative ethics that events that occur before a person is born can be intrinsically bad for the person. But FA4 is incompatible with this. Consider some pre-vital harm, x, that occurs at t_0 and that is intrinsically bad for S, who begins existing at t_1. FA4 implies that x gives S reason to be averse to x *before or during* t_0. But this is impossible; nothing can give a reason to someone who doesn't exist. Since it can't be that x gives S a reason at t_0, FA4 is incompatible with x's being bad for S.

Note that this objection applies to FA3, the other temporally-indexed theory, too. The time at which x, the pre-vital harm, is intrinsically bad for S would seem to be t_0, the time at which x occurs (what other time would it be?). But then FA3 will likewise imply the contradictory thought that x gives S a reason to have an attitude at a time at which S doesn't exist.

Even if there is no such thing as pre-vital harm, non-experienced, non-temporally biased evils make trouble for FA4 in a more abstract way as well. The fact that some evils merit a bias towards the future while other evils do not makes FA4 seems like a bizarre view. There are these evils that provide their victims reason to be against them at every time the victim exists, whether this time is before or after the evil event. In "sending out" the reasons these evils send out, these evils don't discriminate between the past and the future—they send out their reasons in both directions, as it were. But FA4 does discriminate between the past and the future. FA4 in fact says that what *makes* these evils evil is merely the fact that the evils send their reasons into the present and the past. The fact that these evils also send reasons into the future is irrelevant to whether these evil events are evil. That is an odd thing for those attracted to a fitting-attitudes theory to say. Surely the spirit of the fitting-attitudes approach demands the idea that part of what makes my being ridiculed last week bad for me is that it now gives me reason to be against it.

that ignoring the wishes of the dead harms the dead provides a simple and satisfying explanation for why we ought to honor the wishes of the dead.

[25] Brueckner and Fischer (1986: 216) note this point, too. Along similar lines, Parfit observes that "The bias towards the future does not apply to many kinds of event, such as those that give us pride or shame" (1984: 172).

[26] It is undeniable that an event that takes place before a person exists could be *instrumentally* bad for that person.

[27] Cf. Parfit 1984: 495.

4.5. Two Timeless Perspectives

i. Averaging Reasons over Time Normative theories of welfare of the preferentist sort face problems concerning preferences and time as well, for example, the so-called problem of changing desires.[28] Philip Bricker (1980) and Thomas Carson (2000: 86) have introduced ideas that may provide solutions to the problem of changing desires. A strategy shared by each is to construct a sort of "timeless standpoint" (Bricker 1980: 400), built out of all the desires (or rational desires) the subject has at particular times, to arbitrate between his changing desires. Perhaps an idea along these lines could be used by the fitting-attitude theorist. She could propose that what it is for something to be good for someone be explained, not in terms of the subject's reasons at any particular time, but in terms of the average of the strength of the reasons the person has throughout his whole life.[29] This proposal has promise for solving our time-related problems because it offers a fitting-attitude-theoretic way for value to be unchanging. The average of the strength of the reason someone has to want a certain thing over each moment of his life doesn't change.

Call this view 'FA5'. Instead of stating it formally, let's appreciate how it might deliver a more plausible result in a case like Parfit's *Past and Future Operations*. For purposes of illustration we can change the case so that Parfit must undergo *both* the ten-hour and the subsequent one-hour operation. FA5 yields a more acceptable result here than the earlier theories because, if we take as our data facts about what reasons Parfit has at what times, and we "input" these facts into FA5, the "output" FA5 generates is about right.

[28] See Brandt 1982: 179, Bykvist 1998, Carson 2000: 84–87, and Heathwood 2006: 541–2.

[29] I am grateful to Tom Carson for suggesting this application to me.

Even though, at the relatively brief time between the two operations, Parfit has more reason to be against the future suffering than he has to be against the past suffering, this temporary "reversal of reasons" will be swamped by the fact that, at every moment prior to the first operation, Parfit will have a much stronger reason (about ten times as strong) to be averse to the first ordeal, the one that is about ten times worse. So long as there are enough such moments, FA5 will deliver the result that the worse ordeal is about ten times as bad as the lesser ordeal.

But FA5 goes wrong precisely because there might not be enough such moments. Consider what happens if we make the interval between the two episodes of suffering a larger proportion of the whole life in question. If, say, the interval of time between the two operations equals the amount of time Parfit is alive before the first operation, the value of the first, worse episode of suffering will be brought down to only about five times as bad as the lesser episode—and this despite the fact that its intrinsic nature doesn't change between this case and the one above. Worse, if we continue changing the case in this way, making the interval of time between the two episodes almost as long as the whole life in question, the result will be that the second, much briefer ordeal is actually a worse harm than the first, much longer one (the time between the two intervals, during which Parfit has reason to be against the second but not the first, would swamp the time before the first, during which Parfit has reason also to be against the first).

The idea of the average reason someone has throughout his life to want some state of affairs to occur for his own sake is in at least one way more like the idea of the value of that state of affairs for him: both of these quantities are unchanging. But I think it is clear that they are not the same thing.

ii. Counterfactual Analyses The final proposals we'll consider are based around the idea that for something to be good for someone is not for it *actually* to provide her reason to want it but instead for it to be such that it *would* provide her reason to want it, if certain specified conditions held.[30]

Many counterfactual analyses of value specify conditions of *full information*. But this won't help here. Our judgment that Parfit has no reason to be averse to his past pain doesn't depend upon his lacking any information. Were Parfit fully informed, his past pain still wouldn't give him reason to be averse to it.

Another thought is that for something to be good for someone is for it to provide her reason to want to undergo it again, if she could. But this assumes that all goods are things that we undergo, and we have already

[30] Proposals along these lines have been suggested to me, in different ways, by Gunnar Björnsson, Fred Feldman, Pat Greenspan, and Peter Vranas.

discussed many putative examples—non-experienced goods, holistic goods, non-located goods—for which this does not hold.

The problem that time makes for fitting-attitude analyses of welfare stems from the fact that the subjects of welfare, to whose reasons welfare is reduced, are *located in time*. This might suggest that we consider making the relevant counterfactual conditions those in which the subject herself occupies some sort of timeless or atemporal perspective. If there is such a thing as the reasons one would be provided by some event were one to occupy a position outside of time, perhaps these reasons (unlike the actual reasons provided to us as we actually are in time) will be stable enough to provide a plausible grounding for value.

So consider

FA6 x is intrinsically good for S to degree n iff if S were to occupy an atemporal perspective, x itself would give S, while in this atemporal location, reason of strength n to intrinsically desire x for S's sake;

x is intrinsically bad for S to degree n iff if S were to occupy an atemporal perspective, x itself would give S, while in this atemporal location, reason of strength n to be intrinsically averse to x for S's sake.

According to FA6, when some event—an ordinary event that occurs in time, such as someone's being in pain at some time—is bad for some person, its being bad for this person consists in the following fact: if this person, who is actually located in time, were to be located outside of time, the event would, under these circumstances, give the person reason to be against it for his own sake.

FA6 is pretty wild, but that's not my main problem with it (though it may indeed be a problem[31]). It is rather just that, despite its extravagance, FA6 doesn't seem to help. Suppose Parfit were to occupy an atemporal location. From this perspective, he considers some episode of suffering his actual life contains (from this perspective, it is a merely counterfactual episode). Should he be averse to this episode for his own sake?

To try to answer this, I want to ask, Is it true of atemporal Parfit that, despite "currently" being in this atemporal location, he nevertheless *will undergo* the episode of suffering under consideration (as if he is looking down from eternity on the life he is about to begin)? If the answer is Yes, then I think he does have reason to be averse to it (and so FA6's verdict would be correct). But if this is how we understand FA6, then this theory really

[31] Since the concept of an atemporal location at which we could exist might be incoherent. It might be incoherent because it's impossible for time not to exist. Or it might be incoherent because, even it's possible for time not to exist, it would not have been possible for any of us to exist, had time not existed.

just amounts to the tensed theory, FA4, considered above, and inherits its defects. (I am ignoring the problems with the apparently incoherent idea that it can be true of atemporal beings that certain things *will* happen to them.)

So suppose instead that if Parfit were in this atemporal location, then it would neither be true that he will undergo nor true that he did undergo this episode of suffering (the most that is true is that he would have undergone the episode, had he been a temporal being). If this is how it is, then I think it is not at all clear what attitudes Parfit ought to have. We would essentially be asking atemporal Parfit this: suppose you, atemporal Parfit, were a temporal being and were to undergo such-and-such episode of suffering; how do you, as you are, feel about this merely possible episode of suffering, one that you would have underdone, had you been a temporal being?

I'm inclined to think that, just as it is reasonable not to care about one's past suffering, it is reasonable not to care about some episode of suffering that one knows is merely possible for one.[32] Moreover, even if we ought to care at least a little bit about some suffering we would have undergone in some counterfactual situation, surely we don't have as much reason to care about such suffering as we do about our actual future suffering (just as, as already discussed, even if we ought to care at least a little bit about our past suffering, we don't have as much reason to care about such suffering as we do about our actual future suffering).

There is an interesting complication to consider. Should we be inquiring into the reasons atemporal Parfit has to be averse to the suffering in question for the sake of atemporal Parfit, or for the sake of Parfit as he actually is? Either option, it seems, is unsatisfactory. The first option seems unsatisfactory, for why should atemporal Parfit feel bad *for his own sake* about some pain he himself never in fact undergoes? And the second option seems unsatisfactory, too. If actual Parfit still has to undergo the suffering in question, then it seems that, were he atemporal, he would have reason to want that suffering not to occur for his actual self's sake. And if actual Parfit already underwent the suffering in question, then it seems that, were he atemporal, he would not have reason not to want that suffering to occur for his actual self's sake. It's over and done with, after all. We are back again at the original problem.

I think that whatever plausibility FA6 might seem to have is gotten in an illegitimate, question-begging way. We consider some event that is bad for someone. We ask whether the person would have reason to be against it for his own sake if the person were to consider the event from an atemporal point of view. I think, at least initially, we don't really know what to say about the reasons we would have in an atemporal location about the actual events of our lives; but we want to be accommodating and answer the question, so we infer that, since the event we're asked to consider is bad, we must have reason to be against it. But this inference makes use of a simplistic fitting-attitude view of welfare, the very theory under dispute. Once we appreciate, from considering the failure of past pain to provide reasons, that this simple inference is fallacious, we should refrain from using it in the atemporal case, too. And once we so refrain, we return to having no clear idea about what reasons we'd have concerning this bad event if we failed to be located in time. And then, finally, when we make efforts to overcome this puzzlement, I think it becomes clear enough that, were we to be atemporal, we'd have, at best, just a little reason to be against the bad event—certainly not as much reason as we have to be against our actual future suffering.

[32] Note that I am not claiming the following: that we don't have reason to care about some future merely possible episode of suffering that nevertheless is, for us, *epistemically possible* (that is, is an episode of suffering that, for all we know, will actually happen).

Certain things are good for us and certain things are bad for us. We also often have reason to want certain things to occur, or to have occurred, for our own sakes. Similarly, certain things are good and bad *simpliciter*, and we often have reason to want certain things to occur, or to have occurred. Surely these notions of value and of reasons have something to do with one another. But the connections are untidy, and, in any case, less tidy than any attempt to reduce one to the other requires.

REFERENCES

Anderson, Elizabeth (1993) *Value in Ethics and Economics* (Cambridge, MA: Harvard University Press).

Baldwin, Thomas (2006) 'Keynes and Ethics' in R. E. Backhouse and B. W. Bateman (eds.), *The Cambridge Companion to Keynes* (Cambridge: Cambridge University Press), 237–56.

Brandt, Richard (1946) 'Moral Valuation' *Ethics* 56: 106–21.

____ (1959) *Ethical Theory* (Englewood Cliffs, NJ: Prentice-Hall).

____ (1982) 'Two Concepts of Utility', in H. B. Miller and W. H. Williams (eds.), *The Limits of Utilitarianism* (Minneapolis: University of Minnesota Press), 169–85.

Brentano, Franz (1889) *The Origin of Our Knowledge of Right and Wrong*, tr. Roderick M. Chisholm (London: Routledge & Kegan Paul, 1969).

Bricker, Philip (1980) 'Prudence' *Journal of Philosophy* 77: 381–401.

Broad, C. D. (1930) *Five Types of Ethical Theory* (London: Routledge & Kegan Paul).

Brueckner, Anthony L., and Fischer, John Martin (1986) 'Why Is Death Bad?' *Philosophical Studies* 50: 213–21.

Bykvist, Krister (1998) *Changing Preferences: A Study in Preferentialism*, doctoral dissertation (Uppsala: Acta Universitatis Uppsaliensis).
Carson, Thomas L. (2000) *Value and the Good Life* (Notre Dame, In: University of Notre Dame Press).
Chisholm, Roderick M. (1986) *Brentano and Intrinsic Value* (Cambridge: Cambridge University Press).
Danielsson, Sven, and Olson, Jonas (2007) 'Brentano and the Buck-Passers' *Mind* 116: 511–22.
D'Arms, Justin, and Jacobson, Daniel (2000) 'Sentiment and Value' *Ethics* 110: 722–48.
Darwall, Stephen (2002) *Welfare and Rational Care* (Princeton: Princeton University Press).
Ewing, A. C. (1947) *The Definition of Good* (London: Macmillan).
Gibbard, Allan (1992) 'Reply to Blackburn, Carson, Hill, and Railton' *Philosophy and Phenomenological Research*, 42: 969–80.
Heathwood, Chris (2003) review of Stephen Darwall, *Welfare and Rational Care, Australasian Journal of Philosophy* 81: 615–17.
____ (2006) 'Desire Satisfactionism and Hedonism' *Philosophical Studies* 128: 539–63.
Kagan, Shelly (1998) *Normative Ethics* (Boulder, CO: Westview Press).
Keynes, John Maynard (1905) 'Miscellanea Ethica', unpublished.
Kraut, Richard (1994) 'Desire and the Human Good' *Proceedings and Addresses of the American Philosophical Association* 68: 39–54.
Lemos, Noah M. (1994) *Intrinsic Value: Concept and Warrant* (Cambridge: Cambridge University Press).
McDowell, John (1985) 'Values and Secondary Qualities', in T. Honderich (ed.), *Morality and Objectivity: A Tribute to J. L. Mackie* (London: Routledge & Kegan Paul), 110–29.
Moore, G. E. (1903*a*) *Principia Ethica* (Cambridge: Cambridge University Press).
____ (1903*b*) 'Review of Franz Brentano: *The Origin of Our Knowledge of Right and Wrong*' *International Journal of Ethics* 14: 115–23.
Nagel, Thomas (1970) 'Death' *Noûs* 4: 73–80.
Olson, Jonas (2004) 'Buck-Passing and the Wrong Kind of Reasons', *Philosophical Quarterly* 54: 295–300.
Parfit, Derek (1984) *Reasons and Persons* (Oxford: Oxford University Press).
____ (2001) 'Rationality and Reasons', in D. Egonsson, B. Petersson, J. Josefsson, and T. Rønnow-Rasmussen (eds.), *Exploring Practical Philosophy: From Action to Values* (Aldershot: Ashgate), 17–41.
Rabinowicz, Wlodek, and Rønnow-Rasmussen, Toni (2004) 'The Strike of the Demon: On Fitting Pro-Attitudes and Value' *Ethics* 114: 391–424.
____ ____ (2006) 'Buck-Passing and the Right Kind of Reasons' *Philosophical Quarterly* 56: 114–20.
Rønnow-Rasmussen, Toni (forthcoming) 'Analysing Personal Value' *Journal of Ethics*.

Scanlon, T. M. (1998) *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
Sidgwick, Henry (1884) *The Methods of Ethics*, 3rd edn (London: Macmillan).
____ (1907) *The Methods of Ethics*, 7th edn. (London: Macmillan).
Slote, Michael (1983) *Goods and Virtues* (Oxford: Oxford University Press).
Stratton-Lake, Philip (2005) 'How to Deal with Evil Demons: Comment on Rabinowicz & Rønnow-Rasmussen' *Ethics* 115: 788–98.
Stratton-Lake, Philip, and Hooker, Brad (2005) 'Scanlon versus Moore on Goodness', in T. Horgan and M. Timmons (eds.), *Metaethics after Moore* (Oxford: Oxford University Press).
Suikkanen, Jussi (2004) 'Reasons and Value: In Defence of the Buck-Passing Account' *Ethical Theory and Moral Practice* 7: 513–35.
Urmson, J. O. (1968) *The Emotive Theory of Ethics* (Oxford: Oxford University Press).
Velleman, David (1991) 'Well-Being and Time' *Pacific Philosophical Quarterly* 72: 48–77.
Wiggins, David (1987) 'A Sensible Subjectivism?' in D. Wiggins, *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford: Blackwell).
Zimmerman, Michael J. (2001) *The Nature of Intrinsic Value* (Lanham, MD: Rowman & Littlefield).
____ (2004) 'Intrinsic vs. Extrinsic Value', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2004 edn), URL = <http://plato.stanford.edu/archives/fall2004/entries/value-intrinsic-extrinsic/>.