



# LUND UNIVERSITY

## The Strike of the Demon: On Fitting Pro-attitudes

Rabinowicz, Wlodek; Rønnow-Rasmussen, Toni

*Published in:*  
Ethics

2004

[Link to publication](#)

*Citation for published version (APA):*

Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The Strike of the Demon: On Fitting Pro-attitudes. *Ethics*, 114(3), 391-423.

*Total number of authors:*

2

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

## The Strike of the Demon: On Fitting Pro-attitudes and Value\*

*Wlodek Rabinowicz and  
Toni Rønnow-Rasmussen*

According to an influential tradition in value analysis, to be valuable is to be a fitting object of a pro-attitude. If it is fitting to favor an object for its own sake, then, on this view, the object has final value, that is, it is valuable for its own sake. If it is fitting to have a pro-attitude toward an object for the sake of its effects, then its value is instrumental. And so on. Disvalue is connected in an analogous way to contra-attitudes instead.

Apart from the linkage between value and attitudes, what is distinctive for this kind of analysis, at least on some of its readings, is that it establishes a connection between the axiological and the deontic notions: value on this approach is explicated in terms of the stance that should be taken toward the object. That it is fitting to have a certain attitude, that there are reasons to have it, or that the attitude in question is appropriate or called for, are different ways to express this deontic claim. Consequently, an important advantage of the “fitting-attitudes” analysis, or the *FA analysis* for short, is that it removes the air of mystery from the normative ‘compellingness’ of values. There is nothing strange

\* Earlier versions of this article were presented at philosophy departments in Lund and Gothenburg, at a workshop on reasons in St. Andrews, at the Research School of Social Sciences in Canberra, at the fourth European Congress of Analytic Philosophy in Lund, and at a moral philosophy seminar in Oxford. For helpful comments we are indebted to John Broome, Johan Brännmark, Krister Bykvist, Roger Crisp, Jonathan Dancy, Sven Danielsson, Stephen Darwall, Dan Egonsson, Jeanette Emt, Mats Johansson, Jonas Olson, Kevin Mulligan, Ingmar Persson, Björn Petersson, Tim Scanlon, John Skorupski, Michael Smith, Daniel Svensson, Christine Tappolet, Folke Tersman, and Michael Zimmerman. Special thanks are due to the referees, Justin D’Arms and Mark LeBar, for their incisive criticisms and useful suggestions. Our work on this article was supported by a generous grant from the Bank of Sweden Tercentenary Foundation.

*Ethics* 114 (April 2004): 391–423

© 2004 by The University of Chicago. All rights reserved. 0014-1704/2004/11403-0001\$10.00

in the prescriptive implications of value ascriptions if value is explicated in deontic terms.

On the other hand, this deontic format of value analysis invites a natural objection: intuitively, the connection between required attitudes and values is not straightforward. In particular, it appears that in some situations we might well have reasons to have pro-attitudes toward objects that are not valuable. Or vice versa: we might have reasons not to have pro-attitudes toward some valuable objects. The fit between the deontic and the axiological is imperfect, which makes the analysis problematic. Justin D'Arms and Daniel Jacobson, who have identified and discussed this difficulty, call it "the Conflation Problem."<sup>1</sup> Their focus is on what they call "neo-sentimentalist" views, that is, the versions of the FA analysis according to which the attitudes to which values are essentially linked are all forms of sentiment or emotion. Here is the example they use to introduce the problem:

Imagine that you have a rich and generous but touchy friend. . . . If he suspects you of envying his possessions, he will curtail his largesse. That is a good reason not to envy him, if there is any chance whatsoever that you will betray your attitude in action or nondeliberate behavior, but surely it does not speak to whether his possessions are enviable. Another reason you might think it inappropriate to envy him would be based on moral qualms about being pained at a friend's good fortune, but this too seems irrelevant to the ascription of the [evaluative]  $\Phi$  property [= being enviable]. While such good strategic and moral reasons can count in favor of (or against) feeling some sentiment, they seem like the wrong kind of endorsement or criticisms of it. . . . Only certain good reasons for or against having a response bear on the associated evaluative judgment.<sup>2</sup>

In cases like this, there is a potential danger of conflating reasons that do not bear on the evaluation with those that do. Thus, one might be tempted to deny that the friend's good fortune is enviable, if envy would be imprudent or immoral. But, quite independently of this danger of conflation, cases like this present a serious analytical problem. It is imperative for the adherents of the FA analysis to provide a general and noncircular method of distinguishing the reasons that bear on evalua-

1. Justin D'Arms and Daniel Jacobson, "Sentiment and Value," *Ethics* 110 (2000): 722–48. Compare also D'Arms and Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," *Philosophy and Phenomenological Research* 61 (2000): 65–90, and "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)," in *Philosophy and the Emotions*, ed. Anthony Hatzimoysis (Cambridge: Cambridge University Press, 2003). See also Roger Crisp, review of *Value . . . and What Follows*, by Joel Kupperman, *Philosophy* 75 (2000): 458–62.

2. D'Arms and Jacobson, "Sentiment and Value," p. 731.

tion from those that don't. In this article, we sometimes refer to this difficulty as the problem of conflation, but the name we prefer is the "Wrong Kind of Reasons"—problem, or the *WKR problem* for short. We use the latter label to stress that some reasons for pro-attitudes may well be satisfactory but still wrong from the point of view of FA analysis: they may not be the kind of reasons that bear on evaluations.

D'Arms and Jacobson examine in their paper various versions of neo-sentimentalism. Their conclusion is that many neo-sentimentalists have been unaware of the conflation problem and that the few who have been sensitive to this issue haven't dealt with it in a satisfactory way. In our article, we approach the matter in a somewhat different way. Unlike D'Arms and Jacobson, we do not so much scrutinize the extant proposals as try to devise possible solutions to the problem at hand and examine their potential. Furthermore, we consider a broader range of FA theories, not just those of the sentimentalist variety, and we trace back that format of analysis to classical sources. As we show in the brief historical section, the FA approach goes at least as far back as to Brentano and counts among its most important advocates such philosophers as Ewing and Scanlon. Then we briefly take up some problems that face this form of analysis in its pluralist version—the version we ourselves find most plausible. On the monistic versions of this approach, there is just one type of a pro-attitude that is supposed to be a fitting response to all valuable objects, an attitude such as desire, preference, or approval. On the pluralist view, on the other hand, different types of pro-responses fit different kinds of valuable objects. The implication is that values may be fundamentally heterogeneous, since the responses their bearers call for may be quite different.<sup>3</sup>

After this interlude, we move on to the WKR problem. We start with Derek Parfit's distinction between "state-given" and "object-given" reasons for attitudes, which one might try to employ in order to explain what makes some reasons for pro-attitudes inappropriate from the point of view of the FA analysis. We argue that, despite its promise, this solution

3. In fact, on this pluralist view, value bearers themselves can be found in different ontological categories: value can accrue to such diverse entities as states of affairs, things, and persons. Admirable persons call for admiration, desirable states are fit to be desired, and so on. To make things more complicated, one and the same valuable object may well call for a variety of different kinds of pro-responses. Thus, a piece of antique Chinese porcelain may deserve to be admired but it may also call for protection and care. Such conglomerates of fitting responses will typically vary as one moves from one type of valuable objects to another. For an extended defense of value pluralism along these lines, see Elisabeth Anderson, *Value in Ethics and Economics* (Cambridge, Mass.: Harvard University Press, 1993). Compare also Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "A Distinction in Value: Intrinsic and for Its Own Sake," *Proceedings of the Aristotelian Society* 100 (2000): 33–49, and "Tropic of Value," *Philosophy and Phenomenological Research* 66 (2003): 389–403.

does not get us very far. A discussion of other promising but ultimately unsatisfactory solutions leads us to a proposal that we find especially attractive: the idea that the right kind of reasons must play a dual role. Along with justifying pro-attitudes they must also appear within those attitudes as their grounds. In most cases, the proposal works reasonably well, but in the end it turns out that there are special cases in which it seems to fail. There is a way around this difficulty, but—as we argue—it comes at a price: the FA format of analysis loses in generality. The upshot of the article is, therefore, somewhat discouraging. It is not clear whether the conflation problem can be solved after all.

#### HISTORICAL BACKGROUND

We haven't made any thorough study of the historical roots of the FA analysis, but it appears that Franz Brentano may have been the first to put it forward: "We call a thing *good* when the love relating to it is correct. In the broadest sense of the term, the good is that which is worthy of love, that which can be loved with a love that is correct."<sup>4</sup> Love in the relevant sense is a "higher mode" of taking pleasure in something. As such, it is neither compulsive nor instinctive and consists in "the natural feeling of pleasure . . . that is experienced as being correct."<sup>5</sup> The cognitive element in this kind of a pro-attitude—the experience of its correctness that is a part of the attitude itself—is important for Brentano: insofar as they are "experienced as correct [als richtig charakterisiert],"<sup>6</sup> such attitudes can function as the basis of our knowledge of the good. The suggested analysis applies both to what is good in itself and what is good in virtue of something else: in the former case, the object is "pleasing in itself" and in the latter it is "pleasing in virtue of what it brings about or preserves or makes probable."<sup>7</sup> Furthermore, the analysis is extendable to value comparisons: "One might take the better to be that which is worthy of a *greater* love,"<sup>8</sup> but "greater" should not be understood in terms of the greater intensity of feeling. Instead, "it refers to the peculiar phenomenon to be found within the sphere of emotions—namely, to the phenomenon of *preferring*. . . . When we call one good 'better' than another, we mean that the one good is preferable to the other. In other words, it is *correct to prefer* the one good, for its

4. Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, trans. Roderick Chisholm (1889; reprint, London: Routledge & Kegan Paul, 1969), p. 18.

5. *Ibid.*, pp. 21, 22.

6. Chisholm's translation deviates from the German original but may still be faithful to it in spirit. On this point, see J. N. Findlay, *Axiological Ethics* (London: Macmillan St. Martin's Press, 1970), pp. 21–22.

7. Brentano, p. 18.

8. *Ibid.*, p. 25.

own sake, to the other.”<sup>9</sup> Thus, Brentano’s “worthy of greater love” can also be rendered as ‘fitting to be preferred’. But it is noteworthy that he regards preference as a species of emotion, rather than as a purely conative state.<sup>10</sup>

The FA analysis makes a brief reappearance in the writings of C. D. Broad, who presents it in his characteristically cautious way: “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.”<sup>11</sup>

David Ross, on the other hand, opposes the idea.<sup>12</sup> On his view, statements of the form “*a* is good” can be paraphrased as saying that *a* is a “worthy” or “fit” “object of admiration,”<sup>13</sup> but such paraphrases cannot be construed as an outright analysis of ‘good’: “Admiration is not a mere emotion; it is an emotion accompanied by the thought that that which is admired is good. And if we ask on what ground a thing is worthy of being thought to be good, only one answer is possible, namely that it *is* good. It would be absurd to say that a thing is good only in the sense that it is worthy of being thought to be good, for our definition of ‘good’ would then include the very word ‘good’ which we were seeking to define.”<sup>14</sup>

Ross’s reasons for rejecting the FA analysis seem to be twofold. To begin with, the goodness of an object cannot consist in its being worthy of admiration since goodness is the very feature for which the object is being admired. Second, Ross takes the attitude of admiration to involve a judgment that the object admired is good. This would make any analysis of goodness in terms of admiration circular.<sup>15</sup>

The foremost exponent of the FA analysis, A. C. Ewing, rejects in *The Definition of Good* both these objections. He defines “good” as “fitting

9. *Ibid.*, p. 26.

10. Compare Findlay, pp. 25–26. For a modern exponent of the Brentano-inspired approach to FA analysis, see Kevin Mulligan, “From Appropriate Emotions to Values,” *Monist* 81 (1998): 161–88.

11. C. D. Broad, *Five Types of Ethical Theory* (London: Routledge & Kegan Paul, 1930), p. 283.

12. W. D. Ross, *Foundations of Ethics* (Oxford: Clarendon, 1939), pp. 276, 278.

13. At least, if *a* is an action. If it instead is an experience of pleasure, then—insofar it is valuable—it is a fitting object of “satisfaction” rather than admiration (*ibid.*).

14. *Ibid.*, p. 279.

15. Whether this circularity would be vicious depends on what the analysis is supposed to achieve. As Justin D’Arms has reminded us, among modern defenders of FA analysis, Wiggins accepts the circularity and argues that, due to its “detour through sentiments,” the analysis nevertheless is informative to some extent; see David Wiggins, “A Sensible Subjectivism?” essay 5 in David Wiggins, *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford: Blackwell, 1987), p. 189. Compare D’Arms and Jacobson, “Sentiment and Value,” pp. 732–33.

object of a pro attitude,"<sup>16</sup> where "pro attitude" is intended to cover "any favourable attitude to something": "choice, desire, liking, pursuit, approval, admiration."<sup>17</sup> Ewing stresses that different pro-attitudes or conglomerates of such attitudes fit different kinds of valuable objects, which he takes to show that 'good' can have different senses.<sup>18</sup> In response to Ross's objections, he argues that

the reason why it is proper to admire anything must be constituted by the qualities which make the object of admiration good, but it does not follow that the thought that it is good must, if the admiration is to be justifiable, intervene between the perception of the factual qualities admired and the feeling of admiration.<sup>19</sup>

The ground [for a pro-attitude] lies not in . . . goodness, but in the concrete, factual characteristics of what we pronounce good. Certain characteristics are such that the fitting response to what possesses them is a pro attitude, and that is all there is to it.<sup>20</sup>

Thus, (i) admiration is a response to the object's 'good-making' qualities and not to its goodness. For this reason, (ii) the attitude of admiration need not involve or presuppose any judgment that the object admired is good.<sup>21</sup>

Even if Ross's circularity objection is avoidable, FA analysis is threatened by another circle. Terms such as "fitting" or "worthy" invite an evaluative reading. But on that reading, we get a new circularity:  $x$  is valuable = it is valuable to take a pro-attitude toward  $x$ . This circle is avoided by Ewing, according to whom that an attitude is fitting means that it is an attitude one ought to take. In contrast to Moore, it is the deontic notion that is primitive according to Ewing, not the evaluative

16. A. C. Ewing, *The Definition of Good* (London: Macmillan, 1947), p. 152. Among adherents of the FA analysis, one should also mention Richard Brandt, who in "Moral Valuation," *Ethics* 56 (1946): 106–21, suggested that "X is Y-able" (admirable, contemptible, etc.) "means that X is a fitting (or suitable) object of Y-attitude" (p. 113).

17. Ewing, *The Definition of Good*, p. 149.

18. *Ibid.*, pp. 166–67.

19. *Ibid.*, p. 158.

20. *Ibid.*, p. 172.

21. Ewing is not alone in his view that pro-attitudes need not involve any evaluations. In "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)," D'Arms and Jacobson present a critique of various forms of judgmentalism and quasijudgmentalism in treatment of emotions. In their view, emotions such as admiration, envy, etc., do not essentially involve any judgment of value, nor do they even require the agent to entertain any evaluative thoughts. Evaluative language may be highly useful for describing emotions, but the evaluative concepts that are employed in such descriptions need not correspond to any "phenomenological" reality in the emotion itself. Consequently, like Ewing, D'Arms and Jacobson reject the Ross-style circularity objection to FA analysis.

one.<sup>22</sup> The relevant ‘ought’, however, is not the ought of moral obligation. For example, it is not obvious that we are morally required to have a pro-attitude toward pleasure, even though pleasure is a thing of value. Ewing argues that there are two primitive deontic concepts that need to be distinguished from each other: the ought of fittingness and the moral ought. It is the former that should be employed in the FA analysis.

In his later work, *Second Thoughts in Moral Philosophy*, from 1959, he modifies his view in several respects. The ought of fittingness is now interpreted as the ought of “reasonableness”.<sup>23</sup> To say that a pro-attitude is fitting with regard to an object means that the object justifies that attitude or provides reasons for it.<sup>24</sup> Second, he suggests that the relevant pro-attitudes should be understood as primarily conations rather than emotions.<sup>25</sup> Third, he explains that his earlier resistance against framing the FA analysis in terms of the moral ought partly had to do with the fact that emotions are largely involuntary.<sup>26</sup> Conations are more controllable, in his view, and thus amenable to moral requirements. Ewing’s new position is that some senses of ‘good’ should be analyzed in terms of the moral ought while for other senses we need the ought of reasonableness.<sup>27</sup>

After Ewing, FA analysis seems to have disappeared from discussion for several decades. Wittgenstein’s disparaging attitude might have contributed to this. Here is what O. K. Bouwsma recounts from Wittgenstein’s visit to Cornell in 1949:

Towards the end of our discussion which had lasted several hours, W. spoke of A. C. Ewing’s definition—in a Moral Science [Club]

22. Compare Jonathan Dancy, “Should We Pass the Buck?” in *Philosophy, the Good, the True and the Beautiful*, ed. Anthony O’Hear (Cambridge: Cambridge University Press, 2000), pp. 159–75: “Ewing seems . . . to be in a position to say that goodness is not a distinctive property of objects, one . . . to which we should respond with approval and admiration. The goodness of the object just is the relational fact that we should respond to it with approval, admiration or other pro-attitude. The evaluative ‘good’ has been defined in terms of the deontic ‘should’. And with this result, the intuitionists reversed Moore’s position. . . . Moore defined the right, that which we ought to do or should do, in terms of the good. Ewing defined the good in terms of how we should respond” (p. 161).

23. A. C. Ewing, *Second Thoughts in Moral Philosophy* (London: Routledge & Kegan Paul, 1959), pp. 86, 90.

24. To avoid misunderstandings, here and in what follows the notion of a reason is supposed to have an “objective” reading: the existence of reasons for pro-attitudes toward an object is not supposed to be dependent on our beliefs. We might be ignorant about the features of the object that provide reasons for a pro-attitude, but this ignorance on our part does not make such features any less reason-providing.

25. Ewing, *Second Thoughts in Moral Philosophy*, pp. 88–89.

26. *Ibid.*, pp. 92–93.

27. *Ibid.*, pp. 98–100.



lecture, “Good is what it is right to admire.” Then he shook his head over it. The definition throws no light. There are three concepts, all of them vague. Imagine three solid pieces of stone. You pick them up, fit them together and you get now a ball. What you’ve now got tells you something about the three shapes. Now consider you have three balls of or lumps of soft mud or putty—formless. Now you put the three together and mould out of them a ball. Ewing makes a soft ball out of three pieces of mud.<sup>28</sup>

It is amusing to read this witty passage, not least because Wittgenstein’s strict demands of precision that he imposes on definitions seem to be so out of tune with the general tenor of his post-*Tractatus* writings. We believe, by the way, that he is wrong in his critique of Ewing: it is illuminating (if true) to learn that the notions of good, right, and admiration, however vague they might be, are mutually related in the way Ewing’s definition requires. It is informative to learn that the “three pieces of mud” have to be mutually adjusted so as to make up a ball.

In the years that followed, the FA analysis got a short mention in John Rawls’s *A Theory of Justice* (1971)<sup>29</sup> and was then defended by McDowell in 1985,<sup>30</sup> by Roderick Chisholm in 1986,<sup>31</sup> by W. D. Falk in that same year,<sup>32</sup>

28. O. K. Bouwsma, *Wittgenstein: Conversations, 1949–1951* (Indianapolis: Hackett, 1986), pp. 41–42; this passage is also quoted in Mulligan.

29. Rawls provides the following FA analysis of the attributive usage of ‘good’: “A is a good X if and only if A has the properties (to a higher degree than the average or standard X) which it is rational to want in an X, given what X’s are used for, or expected to do, and the like (whichever rider is appropriate)” (John Rawls, *A Theory of Justice* [Cambridge, Mass.: Harvard University Press, 1971], p. 399).

30. According to McDowell, both values and secondary qualities, such as color, sound, etc., are analyzable in terms of certain hypothetical reactions on the part of the subject. However, the two cases are not quite analogous: “The disanalogy, now, is that a virtue (say) is conceived to be not merely such as to elicit the appropriate ‘attitude’ (as a color is merely such as to cause the appropriate experiences), but rather such as to *merit* it” (John McDowell, “Values and Secondary Qualities,” in *Morality and Objectivity*, ed. Ted Honderich [London: Routledge & Kegan Paul, 1985], pp. 110–29, p. 118).

31. On Chisholm’s proposal, A is intrinsically preferable to B if and only if “A and B are necessarily such that, that for any x, the contemplation of just A and B by x requires that x prefer A to B” (Roderick M. Chisholm, *Brentano and Intrinsic Value* [Cambridge: Cambridge University Press, 1986], p. 52).

32. “It has been noted that for ‘x is good’ one may substitute . . . ‘x would be a fitting object of a pro-attitude’, and there are other substitutions: ‘x deserves favor’, ‘x would justify favor’ and ‘there is a case for favoring x’. One may interpret these sentences as expressing the claim that x has a power to evoke favor by way of true comprehension of what it is like. . . . Goodness or value on this showing would be a *dispositional property* of things as truly comprehended, and it would be defined in terms partly psychological and partly not: in terms of power to evoke responses, but responses as they ultimately would be in the ideal case of a *perfect*, no-further-correctible, comprehension of the things in question” (W. D. Falk, “Fact, Value, and Nonnatural Predication,” in his *Ought, Reasons*

by David Wiggins in 1987,<sup>33</sup> by Allan Gibbard in 1990,<sup>34</sup> by Elisabeth Anderson in 1993,<sup>35</sup> and then again by Noah Lemos in 1994.<sup>36</sup> This short list is by no means complete, but not until T. M. Scanlon's *What We Owe to Each Other* from 1998 did the FA analysis experience a true revival in the philosophical community. In his presentation, Scanlon does not refer to the older tradition.<sup>37</sup> Still, there is a clear similarity between Scanlon's version of the analysis and the traditional formulations, even though his proposal is framed in terms of reasons rather than "fittingness" or "oughtness":

[Contrary to Moore, I believe that] being good, or valuable, is not a property that itself provides a reason to respond to a thing in certain ways. Rather, to be good or valuable is to have other properties that constitute such reasons. Since the claim that some property constitutes a reason is a normative claim, this account also

---

*and Morality* [Ithaca, N.Y.: Cornell University Press, 1986], pp. 99–122, pp. 117–18). We are indebted to Stephen Darwall for this reference to Falk.

33. "x is good if and only if x is the sort of thing that calls forth or makes appropriate a certain sentiment of approbation *given the range of propensities that we actually have to respond in this or that way*" (Wiggins, p. 206). This applies not just to "good" but to other value properties as well, for each of which there is "an attitude or response" that is appropriate. Wiggins takes this account of value to be essentially relativistic: "The relativity to us that is here in question consists in the fact that it is we who owe x the response A. . . . What this relativity imports is the possibility that there may be simply no point in urging that a stranger to our associations owes the object this response" (*ibid.*, pp. 202–3).

34. The FA pattern of analysis is advocated by Gibbard for a whole range of moral and evaluative notions. "All these notions concern the way it makes sense to feel about things people do, the feelings that are warranted" (Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* [Oxford: Clarendon, 1990], p. 51). See also his "Preference and Preferability," in *Preferences, Perspectives in Analytical Philosophy*, ed. Christoph Fehige and Ulla Wessels (Berlin: de Gruyter, 1998), p. 241: "To be desirable, we might say, is to be desired fittingly, or justifiably, or rationally. Or since a desirable thing might not be desired at all, we should speak hypothetically: something is desirable if it *would* be reasonable to desire it. It is desirable if desiring it would be *warranted*, if it would *make sense* to desire it, if a desire for it would be *fitting* or *rational*. Likewise, the preferable thing is the one it would be rational to prefer."

35. "To experience something as good is to be favourably aroused by it—to be inspired, attracted, interested, pleased, awed. . . . To value something is to have a complex of positive attitudes toward it. . . . To experience something as valuable and to value it are not to judge that it is valuable. . . . To judge that something is good is to judge that it is properly valued" (Anderson, pp. 1–2).

36. "I suggest that we can explicate intrinsic goodness and badness, and other related value concepts in terms of such concepts as 'being intrinsically worthy of love' and 'being intrinsically worthy of hate'. But what is it for a state of affairs to be intrinsically worthy of love or hate? I propose to follow Chisholm in explicating such concepts in terms of the notion of 'ethical requirement'" (Noah M. Lemos, *Intrinsic Value: Concept and Warrant* [Cambridge: Cambridge University Press, 1994], p. 12).

37. In a footnote, though, he mentions Rawls and refers to a short passage in Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon, 1989), p. 60 (cf. T. M. Scanlon, *What We Owe to Each Other* [Cambridge, Mass.: Harvard University Press, 1998], p. 384).

[i.e., like Moore's] takes goodness and value to be non-natural properties, namely the purely formal, higher-order properties of having some lower-order properties that provide reasons of the relevant kind. . . . It is not goodness or value itself that provides reasons but rather other properties that do so. For this reason I call it a buck-passing account.<sup>38</sup>

Following Scanlon, the FA analysis is nowadays often referred to as the *buck-passing account* of value.

Note that in the quoted passage, Scanlon analyzes the value of the object in terms of reasons to "respond" to that object in certain ways. In other places, he is more explicit on the nature of these responses. For example, in the same section, he writes: "To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it. Exactly what these reasons are, and what actions and attitudes they support will be different in different cases. They generally include, as a common core, reasons for admiring the thing and for respecting it. . . . To claim that something is *valuable* (or that it is "of value") is to claim that others also have reasons to value it, as you do."<sup>39</sup> Thus, pro-attitudes are central to the responses Scanlon has in mind, but his analysis also brings in actions that are expressive of such attitudes.

#### PRO-ATTITUDES?

The virtues of the FA analysis are considerable: it demystifies value and explains why we are justified in our concern for valuable objects. The justification is immediately forthcoming if value is nothing but the existence of reasons for such a concern. Another virtue of the FA analysis is its neutrality on the difficult issues concerning the nature of value judgments. Reducing such judgments to judgments about reasons for pro-attitudes does not beg the question against cognitivist or noncognitivist theories of evaluative discourse. It is consistent with both, since the interpretation of judgments about reasons is left open: one can give them a cognitive (i.e., propositional) reading, but one can also interpret them in noncognitive terms, as recommendations, prescriptions, or expressions of approbation. Still, before we embrace this attractive format of analysis, we must consider how well it can answer some natural objections.

The most plausible versions of the FA analysis are pluralist with respect to value: we must allow for different kinds of value in view of the fact that different valuable objects invite different kinds of pro-responses. Ewing and Scanlon are in this pluralist camp, and we have

38. Scanlon, p. 97.

39. *Ibid.*, p. 95.

ourselves endorsed this view.<sup>40</sup> But value pluralism creates an obvious problem for the FA analysts: if being valuable is defined in terms of fitting pro-responses, and if such pro-responses may considerably vary for different kinds of valuable objects, then what is the common denominator for all the responses in the “pro”-category? What makes them “pro,” one might ask.<sup>41</sup> Actually, there are two questions to be answered: (i) What distinguishes the “pro” from the “contra,” the positive responses from the negative ones? We need to draw a line between value and disvalue. But the second question is perhaps even more pressing: (ii) What distinguishes positive and negative responses from those that are neither? How should the line be drawn between the realm of value/disvalue and the rest of the world?

Intuitively, we think that an object must be valuable if there are reasons to admire it but it need not be valuable just because there are reasons to, say, examine it or reflect upon its properties. Why is admiration a pro-attitude but reflection is not? Clearly, we cannot rest satisfied with Ewing’s uninformative statement: “‘Pro attitude’ is intended to cover any favorable attitude to something.”<sup>42</sup> Nor can we retreat to the Rossian view, according to which such an attitude toward an object always involves an evaluative judgment. Such a maneuver would make the FA analysis circular. Equally circular, of course, would be to define positive (negative) responses as those kinds of reactions that are appropriate with respect to valuable (disvaluable) objects.

To solve this problem, one might suggest that there is a common core in all the positive and negative responses; for example, all of them might turn out to contain a common conative or emotive element. Say, all of them might involve a desire component, some preference or liking. But can we count on it? Perhaps no such common element exists, and all we have is just a complicated network of various family resemblances, without clear borderlines. If the latter is true, then the FA analysis would make the notion of value (disvalue) as indefinite as the underlying notion of a positive (negative) response. While such indefiniteness may be regrettable, we might just have to accept it: after all, what is there to say that our notion of value is fully determinate?

If, however, we demand of philosophical analyses (as opposed to mere accounts of linguistic use) that they be framed in precise concepts, then, in view of such indefiniteness, we might have to conclude that the FA approach does not give us a general analysis of value. Instead, it provides us with a recipe for constructing analyses of specific values.

40. See Rabinowicz and Rønnow-Rasmussen, “A Distinction in Value: Intrinsic and for Its Own Sake,” and “Tropic of Value.”

41. We are indebted to Tim Scanlon for raising this issue (in private communication).

42. Ewing, *The Definition of Good*, p. 149.

According to that recipe, each such value can be defined by reference to specific kinds of pro-responses that the objects exhibiting this value call for. For each specific value property, we would then have a separate FA analysis: one for desirability, another for admirableness, yet another for trustworthiness, and so on. One might wonder, of course, how helpful such a general recipe can be, if we cannot say anything definite about what makes attitudes positive or negative.<sup>43</sup> The answer is that the recipe is meant to be addressed to people who already are “prephilosophically” familiar with the values to be analyzed. They know, we can assume, what attitudes are fitting toward the objects that exhibit the values in question. Thus, for example, they know there are reasons to admire the admirable or to desire the desirable. The recipe then tells them to analyze different values in terms of the corresponding attitudes: to analyze admirability (desirability) as possession of properties that provide reasons for admiration (desire), and so on.

#### WRONG AND RIGHT KIND OF REASONS

Whatever would be the best way of dealing with the problem discussed in the previous section, in this article we focus on another objection to the FA analysis: the WKR problem. It seems that the analysis in question fails to provide an adequate condition for value. That there are reasons to have a pro-attitude toward an object does not yet imply that the object in question must be valuable.

In all essentials, this is the conflation problem that has been posed by D’Arms and Jacobson.<sup>44</sup> However, these two authors mostly concentrate on cases in which there are reasons against having a pro-attitude toward an object, despite the fact that the object in question is valuable. In this article, we focus instead on cases in which there are reasons for a pro-attitude but the object of the pro-attitude lacks value. A case of this kind was used by Roger Crisp in his criticism of the definitions of value that identify value with preferability: “Since there can be reasons for preferring things that have nothing to do with their value, definitions in terms of preferability are false. Imagine that an evil demon will inflict a severe pain on me unless I prefer this saucer of mud; that makes the saucer well worth preferring. But it would not be plausible to claim that the saucer of mud’s existence is, in itself, valuable.”<sup>45</sup>

In general terms, this worry could be put as follows: the reason for a pro-attitude toward an object may have to do with the value of that attitude itself and not with the value of the object. Such a reason may exist even if the object of the attitude is devoid of value.

43. We are obliged to Justin D’Arms for pressing this point.

44. D’Arms and Jacobson, “Sentiment and Value.”

45. Crisp, p. 459.

One might try to resist the objection and argue that, in cases like this, the object does have some value, but its value is indirect: after all, its existence is needed if there is to be a pro-attitude toward it, and that attitude is supposed to be valuable in the case at hand. This response can easily be countered, however. To begin with, many types of pro-attitudes (such as hopes, wishes, or desires) do not presuppose the actual existence of their objects. Consequently, the object's existence need not be a prerequisite for the attitude itself. Second, in the case at hand, it is questionable whether the object has any indirect value at all: if the saucer of mud did not exist, we would not be in this unpleasant situation to begin with. Third, and most important, the response in question is beside the point if we focus on the FA analysis of final value. On that analysis, an object is said to be finally valuable if there are reasons to favor it for its own sake. (To simplify the presentation, we use "favor" as a schematic placeholder for different pro-responses that are called for by different kinds of valuable objects.) The objection then involves setting up a case in which no final value accrues to an object but there still are reasons to favor that object for its own sake, because such a pro-attitude has certain beneficial effects.

We do not really need evil demons for cases like this. More realistic examples are provided by the modern versions of hedonism. Since Butler's *Sermons*,<sup>46</sup> it is a common view that a satisfactory form of hedonism cannot require that we should be pure pleasure seekers. Even though a hedonist takes pleasure to be the only final value, he may well recognize that it would be self-defeating to seek nothing but pleasure: we wouldn't maximize pleasure by such a single-minded pursuit. This "paradox of hedonism" entails that, on the hedonist view, there are reasons to favor, for their own sake, various objects that on that very view lack final value. There may be strong reasons, a sophisticated hedonist might say, to strive for, say, knowledge or freedom for their own sake, in spite of the fact that nothing but pleasure can have final value. Now, even though hedonism may well be an implausible axiological view, value analysis should not prejudge the issue by ruling out that view out of hand. Thus, if FA analysis and hedonism do not go together, as the objection seems to show, then this analysis would have to be rejected.

Similar considerations apply to some other axiological proposals. In the above examples, the value of the attitude has been assumed to be purely instrumental, but that is really an irrelevant restriction. Consider axiological views that ascribe final value to certain pro-attitudes

46. Compare Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, in *The Works of Bishop Butler* (1729), vol. 1, ed. J. H. Bernard (London: Macmillan, 1900), preface, sec. 31, and sermon xi, sec. 1.9; see also Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907), pp. 136, 403.

but deny that kind of value to their objects. An example might be a view according to which final value exclusively accrues to mental states. On some versions of such a position, there may well be reasons of a noninstrumental nature for admiring or cherishing various nonmental objects for their own sake, say, such objects as works of art, persons, landscapes, and so on. Pro-attitudes toward such objects may have features that make them valuable for their own sake and not just for the sake of their effects. On such a view, then, the attitudes in question are ascribed final value, even though their objects are assumed to lack value of this kind. Thus, the objection we have to deal with is quite general: the reason to have a pro-attitude toward an object may have to do not with the object's value but with the value of that attitude itself, be that value instrumental or final. Consequently, the object's value cannot just consist in there being reasons to have a pro-attitude toward the object in question.

In fact, it is not just the value of a pro-attitude that creates the problem. In some cases, the reasons for (or against) a pro-attitude toward an object may instead have to do with deontological considerations, which cannot be traced back to values. It may be morally required, for example, to protect and cherish one's children, however unworthy they may be; not necessarily because (or at least not just because) there is any final or instrumental value in such attitudes but because this is what we ought to do as parents. D'Arms and Jacobson, who in their discussion focus on the wrong kind of reasons against emotive pro-attitudes, issue a warning against "the moralistic fallacy": it is a fallacy "to infer, from the claim that it would be wrong or vicious to feel an emotion, that it is therefore unfitting."<sup>47</sup> While rightness or wrongness of a pro-attitude toward an object may have to do with that object's value, there is no immediate inference from the former to the latter.

The obvious strategy to deal with all such objections is to impose some general requirements on the right sort of reasons for having a pro-attitude toward an object. What needs to be shown is that the reasons that appear in the counterexamples, however strong they might be, are all of the wrong kind: they are not the sort of reasons that the FA analysis is after. How, then, is this Wrong Kind of Reasons problem to be dealt with?

#### OBJECT-GIVEN REASONS

A natural requirement to impose, for an adherent of the FA analysis, is that the relevant reasons for a pro-attitude toward an object must be provided by the properties of the object. Only then can the object be

47. D'Arms and Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," p. 69.

said to be valuable in virtue of the properties in question. It is different with those reasons for a pro-attitude that are provided by the properties of that attitude itself. From the existence of such reasons no conclusions can be drawn about the object's value. In the counterexamples above, the relevant reasons have been all attitude- rather than object-given. In Crisp's example, the reason for desiring a saucer of mud is that this desire would shield one from the punishment by the evil demon. In the hedonistic paradox, the reason for favoring other things apart from pleasure, for their own sake, is that such attitudes are pleasure conducive. And in the "mental state" axiologies that ascribe final value to favoring other things apart from mental states, it is intrinsic features of such favorings that provide us with reasons for holding the attitudes in question.

The distinction between the attitude-given and the object-given reasons for a pro-attitude comes from Derek Parfit.<sup>48</sup> Parfit applies this distinction to desires and uses the label "state-given" (i.e., given by the state of desire) for the reasons that are provided by facts about the desire rather than by facts about its object: "Of our reasons to have some desire, some are provided by facts about this desire's object. These reasons we can call *object-given*. We can have such reasons to want some thing either for its own sake, or for the sake of its effects. . . . Other reasons to want some thing are provided by facts, not about what we want, but about *having* this desire. These reasons we can call *state-given*. Such reasons can also be either intrinsic or instrumental."<sup>49</sup>

In that same paper, Parfit endorses Scanlon's "buck-passing account" of value, but he never connects these two strands of thought: his distinction between the two kinds of reasons is not exploited by him to defend Scanlon's proposal against objections. Indeed, in his presentation of the buck-passing account, Parfit never explicitly addresses the WKR problem.

In a sense, the requirement that the right reasons for a pro-attitude should be object- rather than attitude-given might seem counterintuitive. Surely, fittingness is a relation between an attitude and an object that depends on both the properties of the object and the properties of the attitude. Thus, for example, while it is fitting to admire a courageous person, it need not be fitting to like him or to love him. In other words, it is obvious that the right reasons for a pro-attitude in this sense depend on the properties of both the relata: the attitude and the

48. Derek Parfit, "Rationality and Reasons," in *Exploring Practical Philosophy: From Action to Values*, ed. Dan Egonsson, Björn Petersson, Jonas Josefsson, and Toni Rønnow-Rasmussen (Aldershot: Ashgate, 2001), pp. 17–41.

49. Parfit, pp. 21–22.



object.<sup>50</sup> However, this doesn't yet show that the Parfit-style solution to the WKR problem must be wrong. That solution is supposed to work on the assumption that we have already specified the nature of the pro-attitude in question and that we then ask whether there are any reasons to have such a pro-attitude toward a given object. Given this perspective, in which the pro-attitude already is specified, the idea is that the right reasons for the specified attitude are provided by the properties of the object alone.

Does, then, the distinction between the attitude- and the object-given reasons solve the WKR problem? Not really. The distinction depends on a separation between facts about the object and facts about the attitude, or—to put it differently—it depends on a separation between the properties of the object and the properties of the attitude. It is only the former that can provide object-given reasons. However, drawing such a distinction does not yet solve our problem. It is easy to see that for any property *P* of the attitude there is a corresponding property *P'* of the object: If a pro-attitude toward an object *a* would have a property *P*, then, ipso facto, *a* has (or would have, if it existed) the property *P'* of being such that a pro-attitude toward it would have the property *P*. Consequently, to the attitude-given reason, provided by *P*, corresponds the object-given reason, which is provided by *P'*. In exactly the same way, of course, for any property *P* of the object of the attitude there is a corresponding property *P'* of the attitude itself: the property of being such that its object has (or would have) property *P*. Thus, to each object-given reason corresponds an attitude-given reason, and vice versa.

One might object that this argument presupposes that a reason provided by a property of an object must ipso facto be an object-given reason. And this is too hasty. If the relevant property of the object really is based on a certain reason-providing property of the pro-attitude, then the reason it provides is attitude- rather than object-given.

The trouble with this response is that we need to explain just when a reason-providing property of an object “really” is based on a reason-providing property of the attitude. We could readily explain this, of course, by relying on our intuitions about value: if it is the attitude that we take to be valuable, and not the object, then it is the property of that attitude that is basic and not the corresponding property of the object. In such a case, then, the reason provided by the property of an object would still be attitude-given. But this explanation would beg the question in the present context, in which the question to be answered is precisely what kinds of reasons for pro-attitudes toward an object make

50. We are indebted to Justin D'Arms for raising this issue.

that object valuable, as opposed to those reasons that make valuable the attitude itself.

The claim that the reason provided by a property of an object may be, at bottom, attitude-given can however be understood in a different, non-question-begging way. A property of an object that consists in it being such that a pro-attitude toward it would have the property *P* appears to be quite artificial and somewhat unreal. It is a typical ‘Cambridge property’. At least, it comes out as such if one uses Sydney Shoemaker’s causal criterion.<sup>51</sup> To be real, a property needs to have a causal power (or to be itself a causal power), but in the cases under consideration it is the property of the attitude rather than its mirror image in the object that is fit to play a causal role. While Shoemaker’s causal criterion for genuine properties may well be questioned (what about properties of abstract objects, for example?), the main point may still be valid: a property of an object that consists in it being such that something else would have a certain property is quite suspect, to say the least. Therefore, it might be thought that the WKR problem is solved by the requirement that the relevant reasons for pro-attitudes must be provided by genuine (rather than Cambridge-like) properties of the objects.<sup>52</sup>

This is an attractive suggestion, not least because it relies on a plausible intuition that Cambridge properties cannot be value-making features.<sup>53</sup> But there are instances in which this suggestion won’t do: in those cases, the property of an object that mirrors the value-making property of a pro-attitude is a robust feature of the object, even more robust than the corresponding feature of the attitude. But still the reason provided by the property in question intuitively is of a wrong kind. For a case in point, consider the following modification of Crisp’s example: this time the evil demon demands, on pain of a severe punishment, that we admire him rather than the saucer of mud. The demon’s disposition to punish us if we don’t comply provides a strong reason for admiring him, which is of the wrong kind: it does not make the demon admirable in any way. But there is nothing Cambridge-like in

51. Sydney Shoemaker, “Causality and Properties,” in *Time and Cause: Essays Presented to Richard Taylor*, ed. Peter Van Inwagen (Dordrecht: Kluwer, 1980), pp. 109–35.

52. We owe this suggestion to Mark LeBar.

53. Cambridge properties can of course ground value properties that themselves are Cambridge-like, as, say, the property of being five kilometers to the north from an admirable person. This property may supervene on the Cambridge property of being five kilometers to the north of an unselfish person. But while they cannot be expressed in a nonevaluative language, Cambridge-like value properties are not values, of course. At least, an adherent of the FA analysis of value would claim that they are not since an object might possess such a value property without it being the case that it calls for any pro-attitude.

this feature of the demon. If anything, it is more robust than the corresponding property of the attitude.

It seems, then, that we need to look for another solution to the WKR problem. What is it about the reasons for pro-attitudes in our examples that makes them reasons of the wrong kind?<sup>54</sup>

#### INTERNAL OBJECT-PROPERTIES

If we confine ourselves to final values, there is a solution to the WKR problem that gets support from a long tradition in value theory, going back at least to G. E. Moore. As this tradition has it, the final value of an object, that is, its value for its own sake, is context independent. It is the value the object could retain under any circumstances. As such, it must be grounded in the object's internal properties: the properties of the object that are logically independent of its relations to other objects in its environment. To put it differently, the object's value for its own sake is on this view identified with its value in itself—its intrinsic value.<sup>55</sup> Now, *being such that a pro-attitude toward it would shield one from punishment by the evil demon* is obviously not an internal property of an object. It accrues to the saucer of mud (or to the demon itself, in the modified example) in virtue of certain external facts about the way we would be treated if we didn't comply. Similarly, in the paradox of hedonism, the object's being such that its appreciation for its own sake would be conducive to pleasure is again just a relational property of that object. Being relational, the relevant properties cannot make the

54. One might think the answer has to do with the fact that, at least in the evil-demon cases, the subject is forced to have the pro-attitude, by a threat of punishment. The FA analysis, on this proposal, should only accept as right kind of reasons those reasons that we would have under appropriately idealized circumstances. Absence of threats would be a natural component in such an idealization. This way of dealing with the problem at hand is, however, unsatisfactory: in the first place, no threats are present in the hedonist case, and the evil-demon cases could be reformulated: we could replace the threat of punishment with a promise of reward. One might reply that, in the idealized circumstances, such promises should be excluded as well. More generally, the circumstances are ideal only if nothing else hinges on whether the subject has a pro-attitude or not: the attitude has no further effects. However, this will not do either, and not just because envisaging such very hypothetical situations seems to be quite difficult. More importantly, an idealization like this would not manage to exclude all the counterexamples to the FA analysis: pro-attitudes might have value that is based on their intrinsic features rather than just on their external effects (as we have seen was the case according to mental state axiologies).

55. More precisely, on Moore's view, the intrinsic value of an object can be based on two kinds of properties: the internal ones and the ones that are 'internally relational', so to speak. The latter accrue to the object in virtue of its relations to its own parts or constituents. Thus, like the former, they can in principle accrue to the object in all possible contexts in which the object could be placed (cf. G. E. Moore, *Philosophical Studies* [London: Routledge & Kegan Paul, 1922], pp. 261–62, and *Principia Ethica*, rev. ed. [Cambridge: Cambridge University Press, 1993], pp. 26–27).

object intrinsically valuable, for purely conceptual reasons. So we seem to have an answer to our question: the reasons for pro-attitudes are of the right kind only if they are provided by the internal properties of their objects.

One might wonder, of course, whether the claim that some values are intrinsic is compatible with the FA style of value analysis. If an object is valuable only insofar as there are reasons to have pro-attitudes toward it, then, one might argue, all value is relational by its very nature: the value of an object consists in the existence of a deontic relation between that object and the potential attitude holders. This would suggest that no value is intrinsic, since all value is relational.<sup>56</sup>

The issue is complex and its satisfactory treatment would require a discussion of the very notion of a relational property. In one sense, any value on the FA view seems to be a relational property, since it requires external attitudes toward the value bearer. But at the same time, if a given value-property of an object is grounded in its internal features, then that property even on the FA view appears to be context independent: the fitting pro-attitudes toward the object remain fitting as long as its internal features remain the same, however the external context might change. If relationality is identified with context dependence, then the concept of intrinsic value does seem to be compatible with FA analysis.

Does this mean, then, that the proposal under consideration solves our problem? Is it the case that the reasons for pro-attitudes are of the right kind only if they are provided by the internal properties of the objects? One difficulty with this solution is that, as we have seen, it is applicable only to final values. We cannot extend it to instrumental or contributive values. According to such philosophers as Brentano and Ewing, however, the FA format of analysis is meant to be applicable to all kinds of value. *X* is instrumentally valuable if there are reasons to have a pro-attitude toward *X* not for its own sake but for the sake of its effects.<sup>57</sup> *X* is contributively valuable if there are reasons to have a pro-attitude toward it for the sake of the contribution it makes to some larger whole. And so on. Instrumental and contributive values, by their very nature, are supposed to depend on relational (i.e., external) properties of their bearers.

This difficulty could be dealt with if we restricted the FA style analysis to final values alone. Instrumental and contributive values could then be defined in terms of the final ones (as means to final values and

56. Mark LeBar's comments have alerted us to this problem.

57. Compare Brentano, p. 18; and Ewing, *Second Thoughts in Moral Philosophy*, p. 86.

as their parts, respectively).<sup>58</sup> However, another difficulty is more severe. The proposal is based on the long-standing view that all final value must be intrinsic, but that view itself is disputable. It seems that one might well value an object for its own sake, rather than just as a means to something else, on the basis of its external, relational properties. Thus, to give an example, a valuable object such as a Fabergé egg or a China vase deserves a greater appreciation, for its own sake, if it is rare, or if it is unique in its kind. An object's historical importance, its significant connection to historical personalities or to important events, may also function in a similar value-enhancing way. To illustrate, a dress that has belonged to princess Diana may be seen as valuable because of its connection to that iconic figure. For a different kind of example, think of valuing lives. A life useful to others may be valued for its own sake in virtue of that very usefulness. This instrumentality may be thought to make a life not just more valuable as a means but also more valuable as a final goal. Many more examples can be given, but the main point is straightforward: "The inference from 'value for its *own* sake' to 'value *in* itself' is quick but invalid."<sup>59</sup> Thus, it would not be correct to say that relational properties never provide the right kind of reasons for favoring an object for its own sake. Apparently, sometimes they do.<sup>60</sup>

58. For some problems with such a proposal, see Toni Rønnow-Rasmussen, "Instrumental Values—Strong and Weak," *Ethical Theory and Moral Practice* 5 (2002): 23–43.

59. Rabinowicz and Rønnow-Rasmussen, "A Distinction in Value: Intrinsic and for Its Own Sake," p. 49. Compare also Christine M. Korsgaard, "Two Distinctions in Goodness," in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 249–74, reprinted from *Philosophical Review* 42 (1983): 169–95; John O'Neill, "The Varieties of Intrinsic Value," *Monist* 75 (1992): 119–37; Kagan, "Rethinking Intrinsic Value"; Rabinowicz and Rønnow-Rasmussen, "Tropic of Value."

60. One could try to avoid final nonintrinsic values by means of a reductionist maneuver: suppose it could be shown that, in the last analysis, the value of an object is reducible to the value of "propositional" entities, such as states of affairs or facts. To illustrate, what has final value on this view is not Diana's dress but rather the fact that Diana's dress exists or that this dress has belonged to Diana. If it could be established that final value ultimately accrues to facts alone, then the claim that all final value is intrinsic would become quite plausible: the features that make a fact valuable for its own sake may plausibly be assumed to be intrinsic to the fact in question. In the papers referred to above, we have argued against such reduction, but this argument has not been left unchallenged. For some objections, see secs. 3.1 and 3.2 in Michael J. Zimmerman, *The Nature of Intrinsic Value* (Lanham: Rowman & Littlefield, 2001), and "Intrinsic Value and Individual Worth," in Egonsson et al., eds., pp. 123–38. Since Zimmerman adheres to the view that all final value is intrinsic, he can develop an FA-style analysis of final value that avoids the WKR objection: on that analysis, presented in *The Nature of Intrinsic Value*, chap. 4, the right reasons for favoring a state are provided by that state's internal nature. As Zimmerman puts it, the state must be favored "for its being what it is . . . and not for the sake of its relation to some other valuable state" (ibid., p. 91).

## IS THE PROBLEM REAL?

It seems, then, that we must look for another explanation of what is wrong with our counterexamples to the FA analysis. Such an alternative explanation is suggested by Parfit's discussion of the "state-given" (= attitude-given) reasons for desires in the paper to which we have referred above. According to Parfit, "state-given reasons to *have* some desire are better regarded as object-given reasons to *want* to have it, and to try to have it [where the object that provides these reasons is the desire itself]."<sup>61</sup> Several pages later, he clarifies this suggestion:

Consider . . . the view that our desires are rational if they [are believed to] have good effects. [On Parfit's view, "desires are rational if they depend upon beliefs whose truth would give us reasons to have these desires."<sup>62</sup>] This claim conflates object-given and state-given reasons. If we believe that having some desire would have good effects, what that belief makes rational is not that desire itself, but our wanting and trying to have it. Irrational desires may have good effects. Thus, if I knew that I shall be tortured tomorrow, it might be better for me if I wanted to be tortured, since I would then happily look forward to what lies ahead. But this would not make my desire rational. It is irrational to want, for its own sake, to be tortured. The good effects of such a desire might make it rational for me, if I could, to cause myself to have it. But that would be a case of rational irrationality.<sup>63</sup>

Doubtless, Parfit would say the same about reasons for a desire that depend on that desire's intrinsic features, rather than on its consequences. The relevant reasons are reasons for wanting the desire or for trying to have it. They are not reasons to have the desire.

Essentially the same idea is to be found in Gibbard's *Wise Choices, Apt Feelings*, where he discusses a case of Cleopatra's courtier for whom it is disadvantageous to be angry at her "even if she ordered an execution unjustly and it thus 'made sense' to be angry at her." "In the case of the courtier and the queen, . . . it is rational [for the courtier] *to be angry*, but also rational *to want not to be angry*. This pattern applies not only to emotion but also to belief. . . . Rationally *feeling* or *believing* something is distinct from rationally *wanting* to feel or believe it."<sup>64</sup>

If this claim of Gibbard and Parfit is correct, then the WKR problem disappears. In our counterexamples to the FA analysis, the "wrong" reasons for pro-attitudes are supposed to be provided by their effects or by their intrinsic features. But if what Gibbard and Parfit suggest is

61. Parfit, p. 24.

62. *Ibid.*, p. 25.

63. *Ibid.*, p. 27.

64. Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgement*, p. 37.

right, such reasons simply do not exist. The problematic reasons are not reasons for the attitudes in question; instead, they are reasons for wanting them or for trying to have them.<sup>65</sup> A desire for the saucer of mud would be irrational despite the fact that it shields me from punishment by the evil demon. If I had it, this would be a case of “rational irrationality,” Parfit would say. I have reasons to want to have that desire, or to try to have it, and even, we may suppose, to see to it that I have it (if I can), but I have no reasons to have the desire in question. If Gibbard and Parfit are right, then, the WKR problem seems to disappear.

While attractive, this line of thought is unconvincing. To be sure, if a pro-attitude is either instrumentally or finally valuable, we do have reasons to want it or to try to have it. This accords with the FA analysis, according to which we have reasons for various pro-responses toward valuable objects, attitudes included. But why should we deny, as Gibbard and Parfit want us to, that we also have reasons to have these valuable attitudes?

When D’Arms and Jacobson consider this issue,<sup>66</sup> they point out that a position such as Gibbard’s leads to counterintuitive results:<sup>67</sup> On Gibbard’s view, the courtier endorses being angry at the queen (there is in his view a sufficient reason to be angry), and he also endorses wanting not to be angry at her (there is in his view a sufficient reason to want not to be angry). “But if this is really what [he] endorses then, were his will effective (that is, if he felt what he endorsed feeling), he would feel counterproductive [anger] while vainly desiring not to feel it. This result seems deeply unacceptable.”<sup>68</sup>

As far as we can see, Gibbard and Parfit have no argument for their claim, apart from an appeal to intuitions that we do not share. It seems to us that, in Parfit’s example, the agent does have reasons to desire to be tortured. And in Gibbard’s example, the courtier does have reasons not to be angry. Furthermore, Parfit’s reference in this context to the idea of “rational irrationality” seems somewhat misleading. ‘Rational irrationality’ has been discussed by philosophers in connection with cases such as, say, the Toxine Puzzle, in which, on the one hand, I have a reason to intend an action (drinking the toxine) because I would be rewarded for having that intention, but, on the other hand, I have no

65. They are by no means alone in this claim. For example, John Skorupski takes the same view in his unpublished work on reasons. Note also that an analogous claim is often made with respect to *doxastic* attitudes: beneficial effects of a belief can be reasons for wanting to have that belief or for trying to have it, but it is denied that they can be reasons to believe.

66. D’Arms and Jacobson, “Sentiment and Value.”

67. Their example differs from Gibbard’s, but here we paraphrase what they say using Gibbard’s example.

68. Gibbard, p. 744.

reason to act on that intention. Or cases in which I have a reason to have a certain action disposition but no reason to act on the disposition in question. A case in point is the Newcomb Problem in which my self-denying action disposition is beneficial for me: I am rewarded for not being greedy. But acting on this self-denying disposition is not to my advantage. Another well-known example of the same kind is the well-known paradox of deterrence: I have a reason to be disposed to realize my threats when I am challenged, come what may. If I am determined to act in this way, the chances are that my enemies will leave me in peace. But, if I am challenged, acting on my disposition may well be irrational, if implementing the threat is costly. Parfit himself discusses several cases like this.<sup>69</sup> In all these cases, however, the rewards are tied not to the action itself, but to the state that leads to the action—to the intention or to the action disposition. It is in this sense that such cases exhibit rational irrationality: the action is irrational because of its harmfulness but there are reasons for being in a state that leads to the action, since being in that state has benefits. The cases we now consider are different. The reward is not tied to our wanting to have a desire for the saucer of mud or to our trying to have it. It is dependent on the desire itself. Similarly, in the paradox of hedonism, the reward is tied to favoring other things apart from pleasure, for their own sake, and not to wanting to have these pro-attitudes or to trying to have them.

Surely, if we are supposed to have reasons for actions when the actions have useful effects or are valuable for their own sake, why shouldn't we have reasons for attitudes in comparable circumstances? The two cases appear to be perfectly analogous. *Pace* Gibbard and Parfit, it is therefore plausible to conclude that we do have reasons for beneficial attitudes, even when they are directed to objects without value. To be sure, we also have reasons to want to have such attitudes and to try to have them, but this is because we have reasons to have them, in the first place. Consequently, the WKR problem still stands: why are these reasons of the wrong kind, from the point of view of the FA analysis?

In fact, on second thought, even if we were to suppose, for the argument's sake, that Gibbard and Parfit are right and the reasons of the wrong kind are not reasons for having the pro-attitudes toward the objects, this wouldn't really get us off the hook. The WKR problem would disappear but another (structurally analogous) problem would arise in its stead: while we would no longer need to explain what makes some reasons for pro-attitudes wrong, from the point of view of the FA analysis, we would instead need to clarify, without taking the notion of

69. Compare chap. 1, secs. 2–9 (and esp. p. 13, where the label “rational irrationality” is introduced) in Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).



value for granted, what makes something a reason for wanting (or trying) to have a certain attitude toward an object rather than a direct reason for having the attitude in question. Anyway, we believe that the proposal under consideration is not viable: we do have reasons for pro-attitudes if the pro-attitudes are valuable.

#### THE DUAL ROLE OF RIGHT REASONS

Here is a new solution to the WKR problem that we would like to examine in some detail. On the FA analysis, the value of an object consists in there being reasons to favor the object in question. But favoring is not supposed to be a nondiscerning attitude. Rather, the idea is that we are to favor the object on account of some of its properties.<sup>70</sup> These properties, then, appear in the intentional content of the pro-attitude. At the same time, they are supposed to make the object valuable. Consequently, they also provide reasons for favoring the object. Thus, they have a dual role. On the one hand, (i) they appear in the intentional content of favoring as the features on account of which the object is favored. On the other hand, (ii) they justify favoring the object in that way, that is, provide reasons for the pro-attitude in question.

To illustrate, consider a person who is admirable because of her courage. Her courage has then a dual role: it is the property for which ('on account of which') she is to be admired, and at the same time it provides us with a reason for admiration. Or, to take another example, think of a wilderness that is valuable in part because of its being unspoiled. That property, again, has a dual role: we should cherish the wilderness in part on account of its being unspoiled and its being unspoiled provides a reason for cherishing it in this way.<sup>71</sup>

But why should we require, in the first place, that the reasons for a pro-attitude appear in the content of the pro-attitude as the features for which the object is being appreciated? The answer, we believe, has to do with the close connection between such pro-attitudes and value judgments. It should be possible for a person who has this kind of a pro-attitude toward an object to move on to an evaluative judgment of the object, in which the properties that make the object valuable are

70. Compare, e.g., Broad's insistence on the subject of the pro-attitude having "an adequate idea of [the object's] non-ethical characteristics" (Broad, p. 283) or Falk's suggestion that it is a matter of favoring the object "by way of true comprehension of what it is like" (Falk, p. 117).

71. Note that this duality of roles characterizes both the properties that make an object valuable for its own sake and the properties that make it valuable as a means. On our interpretation of the FA analysis, this format of analysis is just as applicable to instrumental value as to final value. The instrumental property that provides a reason for favoring an object as a means should at the same time be the property on account of which we favor that object as a means, if the object is to have an instrumental value.

specified. Such a direct move is possible only if these properties already are specified in the intentional content of the pro-attitude.<sup>72</sup>

This is how things work out with the right kind of reasons—with those reasons that the FA analysis is after. But what about the wrong kind of reasons for favoring? In the paradox of hedonism, we are supposed to have reasons to favor other objects apart from pleasure, for their own sake, because this favoring is pleasure conducive. Still, the existence of such reasons is not meant to imply that these objects have final value. Why? Here is how our proposal can answer this query. Consider the case of friendship. On the hedonist view, we are supposed to care for our friend for his own sake, because *caring for him for his own sake would make us happier*. The italicized property of the person in question is a reason for caring, but are we supposed to care for him on account of that reason? Surely not. Caring for him on account of that property would not make us happier. To promote our happiness we should care for our friend for his kindness, humor, intelligence, the things we have experienced together, and so on. Thus the pleasure-related reasons to care, however strong they may be, are of the wrong kind: they do not imply value. That caring for him would make us happier is a reason for caring but it is not what should be our motive for caring. It is not the feature that is supposed to occur in the intentional content of caring as the ground for that attitude. In other words, the feature in question does not have the required dual role. The reason it provides is of the wrong kind from the point of view of the FA analysis.

The same diagnosis applies to the other counterexamples we have considered. A mental state axiology may imply that we have reasons to favor, for their own sake, other things than mental states, where these reasons are provided by some intrinsic features of such favorings. To illustrate, on such a view, we may have reasons to care for our friends, for their own sake, because such a caring attitude has certain internal properties that make it intrinsically valuable. But, again, these reasons for caring are of the wrong kind: I do not care for a friend on account of the fact that my caring has certain desirable intrinsic features. Similarly, in the case of the evil demon, desiring the saucer of mud for its own sake would shield us from punishment. But again this reason for desire is of the wrong kind. To desire the saucer of mud for its own sake, we can try to appreciate its color, texture, or even its taste,<sup>73</sup> but

72. It is another thing that this move to a value judgment need not always be correct. I might favor  $x$  on account of it being  $P$ , even though  $P$  does not make favoring  $x$  justified; either because  $x$  in fact lacks  $P$  or because  $x$ 's having  $P$  is not an acceptable reason to favor  $x$  in the first place.

73. What features would fit the bill depends on what we are supposed to desire for its own sake in this case: is it that we have the saucer of mud, that we eat the mud, or something else? The example is underdescribed in this respect.

surely it would not do to desire it on account of the consequences of this desire: we would then not be desiring it for its own sake.<sup>74</sup>

There is a certain feature of our proposal that may be seen as problematic. If the idea of the dual role for the right kind of reasons is to be viable, the relevant pro-attitudes must have a relatively complex intentional content: they must consist in favoring an object on account of some of its properties. A favoring that is not discerning in this way will not do for our purposes. This requirement of discernment, however, may be seen as too demanding, since some of our positive responses to objects need not be focused in this way on the particular features of the objects in question. For example, romantic love has often been said to be unconditional: if I love my spouse in this way, I do not love her on account of this or that. Her particular characteristics may indeed cause my love for her; they may also provide reasons for my love, but they are not what I love her for. If our interpretation of the FA analysis is correct, the existence of reasons for such a 'nondiscerning' pro-attitude toward an object would not by itself imply that the object is valuable.

Is this a serious problem for our proposal? Not necessarily. In a case like this one, it may be true that I have reasons to love my spouse in a romantic, "nondiscerning" way. Her characteristics (beauty, charm, her love for me, our common experiences, or what have you) may well provide me with such reasons. But then I should also have reasons to cherish her on account of those characteristics that make her so lovable. Thus, her value to me is not endangered by the requirement that I love her nondiscerningly, as long as she also deserves my discerning appreciation. This point can be generalized, we believe, to all those cases in which certain features make an object valuable. Such features may pro-

74. But suppose the evil demon is even more perverse than we have assumed. Suppose he demands of us, under a threat of punishment, not just that we desire the saucer of mud for its own sake but that we desire it for its own sake precisely on account of it being such that desiring it in this way would shield us from punishment. What then? Under such circumstances, the reason for the desire would at the same time be the feature on account of which we are supposed to desire the object. Would this make it a reason of the right kind? On our interpretation of the FA analysis, it appears we should have to admit this. It would be nice if we could resist this objection and argue that the attitude the perverse demon requires of us is conceptually impossible and thus nothing that we can have reasons for. It certainly does appear to be psychologically impossible to desire something for its own sake on account of the fact that this desire would have beneficial effects. No sane person would be able to have such an attitude. But a psychological impossibility is not the same as a conceptual inconsistency. If the demands of the perverse demon in this imagined example are conceptually coherent, then our interpretation of the FA analysis founders for cases like this. Still, since this counterexample is so contrived in its lack of psychological realism, the difficulty appears to be rather marginal. A less contrived counterexample will, however, be considered in the next section.

vide us with reasons to favor the object in a nondiscerning way, but they will also provide us with reasons to favor it on account of the features in question.

A variant of this idea can be employed to answer another query concerning our proposal. Is the appeal to the requirement of the double role for reasons still adequate when we focus on beliefs rather than pro-attitudes? Even in the case of beliefs we can intuitively distinguish between attitude-given and object-given reasons for belief. Thus, for example, it may be beneficial for us to believe that a serious problem we face in our life has a solution. Even if we have no “objective” grounds for this belief—no evidence for its truth—having that belief may still make our life more bearable. Will the double-role requirement account for the distinction between the two kinds of reasons even in the doxastic case? At first sight, this may be doubted, since beliefs, unlike pro-attitudes, normally are “nondiscerning”: Ordinarily, when we believe that a certain state of affairs obtains, for example, that it is raining outside, the content of this belief does not involve the reasons on account of which we find that state likely or probable, such as that our cat, who has just got in, is soaking wet. So it seems that the reason for our belief, even though it is a reason of the right kind, does not appear in the belief itself as its evidential ground. Still, we can now recognize how this difficulty can be dealt with. The features that provide us with doxastic reasons of the right kind have two roles: they give us reasons to believe that a certain state of affairs obtains but they also provide premises for our inferences to those beliefs. I infer that it is raining from the fact that my cat has got wet. The reasons for beliefs are of the wrong kind if they lack this inferential role. Thus, in the example above, the reason for our belief that the problem we face has a solution is a case in point: this belief is not inferable from the fact it would make my life more bearable.

Before we move any further, let us briefly consider another problem. On the FA analysis, the existence of reasons for pro-attitudes implies value. But the FA analysis also makes the converse claim: value implies the existence of reasons for pro-attitudes. Is this undisputable? Consider the following modification of the evil-demon example: Suppose that *a* is a valuable object but the evil demon forbids us to have any pro-attitudes toward *a*, on pain of severe punishments. Doesn't it then follow that we have no reason to have such attitudes toward *a* in spite of the fact that *a* is valuable? If this were true, the FA analysis would be in trouble.<sup>75</sup>

75. We owe this objection to Jonas Olson. He has also provided us with another example that illustrates this problem. On the strong reading of the paradox of hedonism, hedonism not only implies that we have reasons to pursue other things, apart from plea-

In our view, this objection can be dealt with. In the case described, we do have reasons, of the right kind, to favor the object, but we also have contrary and, let us assume, stronger reasons to abstain from such favoring. The latter reasons are of the wrong kind from the point of view of the FA analysis, but this does not make them any less compelling in the case at hand. Still, the answer to the objection is that these reasons for abstaining do not undercut the right reasons for favoring. The “wrong” reasons for abstaining may override the right reasons for favoring, but they neither nullify the latter nor show them to be spurious.

Needless to say, reasons of the right kind may themselves pull in opposite directions. Some features of an object may provide us with right reasons for favoring, while other features may give us right reasons for abstaining from favoring or even right reasons for disfavoring. But this is, of course, as it should be. Clearly, an object may be valuable in some respects and indifferent or disvaluable in other respects. In some cases, but probably not always, it makes sense to try to aggregate these partial evaluations into one value judgment *tout court*. Whether an object will then be assessed as valuable all things considered will depend on the balance of reasons of the right kind for various attitudes toward the object (favorings, disfavorings, and indifference). How such a balancing is to be done is a difficult issue.<sup>76</sup> But only reasons of the right kind will be relevant for determining the balance.

---

sure, for their own sake, but also that we have reasons not to pursue pleasure itself in this way: striving for pleasure, for its own sake, is self-defeating on this view. Note also that all the examples of the wrong kind of reasons with which D’Arms and Jacobson (see their “Sentiment and Value”) illustrate the conflation problem are of this kind: they concern cases in which we have reasons not to have pro-attitudes toward valuable objects.

76. We are indebted to Sven Danielsson for getting us to think of the difficulties involved in the enterprise of aggregation. For example, can right counterreasons undercut rather than simply override right reasons for favoring? It seems that such undercutting might sometimes be possible: *prima facie*, we have a reason to admire a person for her courage, but this reason is undercut, it seems, if it is the courage of a murderer. It appears, therefore, that whether a certain property of an object does provide a right reason for favoring may depend on other properties of that object. To take another example of an aggregation problem, suppose that one property of an object provides a right reason for favor and another property gives an independent right reason for disfavor. Say, we have a reason to admire a person for her courage but also a reason to despise her for deceitfulness. Assume we come to the conclusion that the former reason overrides the latter: taking both features into account, favoring is more appropriate, on balance. But what sort of favoring are we then talking about? And on account of what features? Is it on account of the conjunctive property courage-and-deceitfulness? But surely, deceitfulness is despicable. So how can we favor her on the account of both features taken together? It seems we should favor her on-account-of-her-courage-and-despite-her-deceitfulness. But this makes the pro-attitudes toward the object even more complex in their intentional content than we have heretofore assumed.

#### THE DEMON STRIKES BACK

For a long time we thought that the dual-role solution would adequately deal with the WKR problem. But Folke Tersman has showed us (in private communication) that the demon has yet another card to play. Recall the modified version of the demon example in which the demon requires us to admire him, rather than a saucer of mud, under a threat of severe punishment. But now suppose that he wants us to admire him, for his own sake, precisely on the account of his determination to punish us if we don't. Under these circumstances, it seems that the dual-role requirement is satisfied: the demon's determination to punish us if we don't comply provides the reason for our admiration and at the same time appears in the intentional content of that attitude as the feature for which its object is being admired.<sup>77</sup>

But is the example psychologically plausible? Isn't fear rather than admiration the natural attitude in this situation? Well, fear is certainly natural in this case, but psychologically it seems quite possible to admire a person for his frightening and malicious set of mind: at least some of us, sometimes, appear to be drawn to evil in this way. Even if one were to argue that we can never admire anyone for the features we take to make him bad,<sup>78</sup> it is enough in the example under consideration to admire the demon on account of his determination to punish us otherwise; we are not required by him to see that feature of his as malicious or bad-making.

We might, of course, bite the bullet and say that, in this new demon example, the reasons for the admiration are of the right kind. We might insist that, in this case, the object of admiration is valuable because the reason for the admiration has the required dual role. However, intuitively, this bullet seems too hard to bite: while it is true that we may be attracted to evil, this does not make evil a good thing. Determination, as such, may be admirable; strength of will is generally regarded as a valuable character trait. But the demon's determination to punish those who do not admire him is not admirable but despicable. And in the example under consideration, it is that latter feature that has the required dual role. The new demon example shows, we believe, that a reason for a pro-attitude may have a dual role but still be the wrong

77. There is some self-reflexivity in the attitude required by the demon: we are to admire him, for his own sake, for his readiness to punish us if we do not admire him for that feature. But this self-reflexivity need not make the required attitude incoherent. Thus, the demon's demand need not be conceptually impossible to fulfill. Nor, as we are going to argue, is there any psychological impossibility in this case. In that respect, Tersman's example differs from the case of a perverse demon who asks us to desire, for its own sake, a lump of mud, and to do it precisely on account of the beneficial effects of that very desire.

78. Mark LeBar has suggested this objection in his comments.

kind of reason from the point of view of the FA analysis. The demon's determination to inflict punishment on us if we don't comply is a feature that makes him worse, not better.

#### ATTITUDINAL CONCERNS

This is bad news for the dual-role proposal. If Tersman's counterexample can be upheld,<sup>79</sup> then the proposal in question does not provide a fully satisfactory solution to the WKR problem. At least, not as it stands. Justin D'Arms has suggested (in his comments to this article) how the proposal could be rescued. Consider an emotional attitude such as shame, for example. For a person to be ashamed, he must view what he is ashamed of as a weakness or disability on his part. Similarly, to be amused at something, a person must see some features of that object as incongruous, in one way or another. Again, to admire someone, for some of his features, we must view these features as strengths or achievements on his part. It seems, then, that different types of pro- and contra-attitudes, by their very nature, exhibit certain characteristic kinds of concerns: different kinds for different attitudes.<sup>80</sup> Now, consider Tersman's demon: we are supposed to admire him for his determination to punish us if we do not admire him in this way. But insofar as we admire him for this feature, then, by the very nature of admiration, we must view that feature as a strength on his part. Now, as we know, the feature in question at the same time justifies our admiration. But it does so not in virtue of being a strength on the demon's part but simply because we have an interest not to be punished. What makes it a reason of the wrong kind is this disparity between the way in which the property justifies the

79. It may not be such a good counterexample after all. Stephen Darwall (in private communication) questions its psychological plausibility: "It is true that it is possible to admire a person for his threatening and frightening way. But that is not the same thing as admiring him because he is threatening *you*. [And it is this latter circumstance that is the precise reason why *you* should admire him.] So, in the case at hand, we are not admiring him, for his own sake, 'precisely on account of his determination to punish us if we don't'. . . . I don't think it is psychologically possible to admire someone for the reason that one would be harmed if one doesn't. . . . And that is so even though it is quite possible to admire someone for being a threatening person." Darwall may be right on this point but what he claims is not obviously true: it is not obvious that people never are able to feel a masochistic admiration. Why should it be psychologically impossible to admire our would-be tormentors at least in part precisely on account of the torment they are determined to inflict upon us if we don't obey their commands? Anyway, even if Darwall is right and the example after all lacks psychological realism, the problem will still stand if such cases at least are conceptually possible.

80. Compare D'Arms and Jacobson, "The Significance of Recalcitrant Emotion."

attitude and the characteristic concerns of the attitude in question.<sup>81</sup> Reasons for an attitude are of the right kind only if they speak to the attitude's characteristic concerns. That is, a property of an object, to be a reason of the right kind, must not only appear in the intentional content of the attitude it justifies: it must justify the attitude in virtue of being the kind of property that the attitude is characteristically concerned with. For example, to be a right kind of reason for shame, a property of mine must justify my shame because it is a weakness on my part, and to be a right kind of reason for amusement a property of an object must justify amusement in virtue of its incongruity. And so on.

Is this solution to the WKR problem workable? Possibly, but it has a number of potential drawbacks. In the first place, concepts such as "strength," "weakness," "incongruity," and so on invite an evaluative reading. But then FA analyses of value that employ concepts of this kind are circular: they presuppose one value concept (say, the concept of strength) in order to analyze another value concept (admirability). The circle is not necessarily vicious: it need not be the case that "strengths" are to be understood as simply those features that make its owner admirable. But in any case, if the notions that refer to characteristic concerns of attitudes themselves are evaluative, then FA analyses of values in terms of such characteristic concerns will never leave the realm of value: they will analyze values in terms of other values. This wouldn't be welcomed by those adherents of FA analysis who see it as a reductive project. Still, this difficulty perhaps can be dealt with: it might be possible to argue that the correct interpretation of the concepts such as "weakness," "strength," and so on is not evaluative and that the characteristic kinds of concerns of different pro- and contra-attitudes can be specified in a descriptive way. If this is the case, then the danger of circularity would be avoided.

Be that as it may, there is another worry about the solution under consideration: it doesn't seem to generalize to all value concepts that FA analysts want to account for. While the idea of a characteristic concern is applicable to many pro- and contra-attitudes, especially those of emotional variety, it doesn't seem applicable to all of them. Consider such values as desirability or preferability, which play a central role in many highly influential axiologies. On the solution under consideration, an object is desirable (preferable) insofar as there are reasons to desire it (to prefer it) that speak to the characteristic concerns of desire (preference). But what characteristic concerns can we ascribe to such conative

81. On second thought, this may have been what Darwall was after in his comments. We can admire the demon for his threatening ways, but the way that feature of the demon justifies our attitude is different from the way it is being admired: our admiration is justified simply because we have an interest not to be harmed.



attitudes? When I desire or prefer a state of affairs, on account of some of its features, is there any constraint, coming from the very nature of the attitudes in question, on the kind of features that can play this role? We doubt it. But if no such characteristic concerns for desires or preferences can be found, then Tersman's demon can come back and demand of us that we desire a certain state of affairs, for its own sake, on account of it being such that we would be punished if we did not desire it in this way.

D'Arms and Jacobson need not take this worry very seriously: their view seems to be that all kinds of values should be analyzed in terms of relatively "thick" attitudes, of emotional variety<sup>82</sup> and that "thin" conative attitudes, such as desire or preference, should be avoided by FA analysts. But such a severe restriction on the FA format of analysis would in our view make this analysis of value much less attractive.

#### CONCLUDING REMARKS

Do our difficulties with the WKR problem show that the FA analysis must be rejected after all? It would be premature to draw this conclusion. Perhaps the FA analysis should be restricted to "thick" attitudes, with a definite focus of characteristic concern. Or perhaps the distinction between the right and the wrong kind of reasons can be clarified in some other way instead. One possibility would be to go back to Parfit's distinction between the attitude- and the object-given reasons. Perhaps, despite our doubts, it is possible to provide a clear division between these two kinds of reasons. However, one should not attach too much hope to this alternative. Even if such a division could be made, it can be questioned whether the object-given reasons for pro-attitudes will always be of the right kind. As we have seen, in those cases when the demon demands to be admired himself, his determination to punish nonadmirers is one of his prominent and robust characteristics. Thus, that feature seems to be a clear-cut object-given reason for our pro-attitude, even though it is a reason of the wrong kind.

Another alternative would be to go back to the founding fathers of the FA analysis—Brentano and Ewing—and make more of the notion of fittingness. The idea would be that this format of analysis rests on a special deontic notion of fittingness (or "correctness," if we use Brentano's terminology), which should not be confused with the standard deontic notions, such as "ought," "should," or "must." In the demon example, it is true that we should have a pro-attitude toward the demon—our reasons for having that attitude may well be quite overwhelming—but it is also true that, intuitively, this pro-attitude is not fitting with respect to its object. A fitting attitude toward someone who

82. Compare D'Arms and Jacobson, "Sentiment and Value."

makes such outrageous demands on us would be hate and contempt, rather than admiration. At least, this is what Ewing and Brentano would certainly insist on. Fittingness, then, may be seen as a special deontic relation that is present in just those cases when the reasons for pro-attitudes are of the right kind.<sup>83</sup>

Intuitively, this sounds right. But as an analytic proposal, it does not seem very promising. For how are we supposed to clarify this basic relation of fittingness? If we leave it as primitive, then FA analyses can be accused of reducing the concept of value to another notion that is at least as troublesome. On the other hand, if we try to explicate this relation, we seem to be drawn to something like this: a pro- or contra-attitude is fitting with regard to its object if and only if it is adequate to the value of that object.<sup>84</sup> Thus, the analysis moves in a circle: from value to fittingness and from fittingness to value.

It may be that this danger of circularity can somehow be dealt with in a satisfactory way. Or there may be some other, better ways of dealing with the WKR problem. But until that problem is fully resolved, the fitting-attitudes analysis of value is not in the clear.

83. One of the editors of this journal has with some justification described our article as an example of “philosophical bait and switch.” On a very natural reading of “fitting” or “worthy,” the WKR problem does not arise for the classical versions of the FA analysis, in which expressions such as these play a crucial role. That problem only emerges when the special notions of fittingness or worthiness are replaced by generic deontic constructions, such as “ought” or “reason.”

84. Unless we are mistaken, this seems to be the position D’Arms and Jacobson take in their “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions.” As they put it, “emotions present things to us as having certain evaluative features. When we ask whether an emotion is fitting, . . . we ask about the correctness of these presentations” (ibid., p. 72). Thus, it is fitting to feel *F* in response to *X*, where *F* is an emotion that represents *X* as having  $\Phi$ , if the evaluative representation *F* contains is correct, i.e., insofar as *X* does have the evaluative property  $\Phi$ . The authors emphasize that the representation in question is not a judgment that *X* has  $\Phi$  but, rather, a form of appearance: in an emotion, *X* appears to have  $\Phi$ . (In their later paper, “The Significance of Recalcitrant Emotion [or, Anti-Quasijudgmentalism],” D’Arms and Jacobson go even further and suggest that that the emotion itself, as distinct from its description, need not involve any evaluative notions.) But the relevant point, in the present context, is that fittingness on this view is supposed to consist in some relation of adequacy between the emotion and the value of its object.