Philos Stud (2015) 172:1455–1474
DOI 10.1007/s11098-014-0359-z

Dissolving the wrong kind of reason problem

Rach Cosker-Rowland

Published online: 31 July 2014

Abstract According to fitting-attitude (FA) accounts of value, *X* is of final value if and only if there are reasons for us to have a certain pro-attitude towards it. FA accounts supposedly face the wrong kind of reason (WKR) problem. The WKR problem is the problem of revising FA accounts to exclude so called wrong kind of reasons. And wrong kind of reasons are reasons for us to have certain pro-attitudes towards things that are not of value. I argue that the WKR problem can be dissolved. I argue that (A) the view that there are wrong kind of reasons for the pro-attitudes that figure in FA accounts conflicts with the conjunction of (B) an extremely plausible and extremely weak connection between normative and motivating reasons and (C) an extremely plausible generality constraint on the reasons for pro-attitudes that figure in FA accounts. I argue that when confronted with this trilemma we should give up (A) rather than (B) or (C) because there is a good explanation of why (A) seems so plausible but is in fact false, but there is no good explanation of why (B) and (C) seem so plausible but are in fact false.

Keywords Wrong kind of reason problem · Buck-passing account of value · Fitting-attitude accounts · Reasons · Value

1 The wrong kind of reason problem

According to fitting-attitude (FA) accounts of goodness and value:

FA. X is non-instrumentally good *simpliciter* or of final value if and only if *X* has features that provide normative reasons for us to have a pro-attitude

R. Cosker-Rowland (✉)
Faculty of Philosophy, University of Oxford, Oxford, UK
e-mail: r.cosker-rowland@leeds.ac.uk

towards *X* (including to admire *X* for its own sake and to desire *X* for its own sake).[1]

In recent years T.M. Scanlon and Derek Parfit are among those who have advocated FA accounts and John McDowell and David Wiggins have embraced the bi-conditional *FA*.[2] And FA accounts have a historical pedigree stretching back through A.C. Ewing to Franz Brentano, and some have suggested, further to Kant.[3] Proponents of FA accounts claim that they provide an intuitive account of final value that explains several connections between reasons and value, is ontologically parsimonious, and demystifies the notion of final value.[4]

The most famous objection to the bi-conditional *FA* is that it overgenerates instances of things that are good *simpliciter* or of final value.[5] If *FA* holds, then if there is a reason to have a pro-attitude towards *X*, *X* is good. But according to this famous objection, there are reasons to have a pro-attitude towards things that are not good. So, *FA* is false. Consider the following two examples:

Desire a Saucer. An evil demon will severely punish us if we do not desire a saucer of mud for its own sake. So, there is a reason for us to desire a saucer of mud for its own sake. But a saucer of mud is not good or of value.

Admire the Demon. An evil demon will severely punish us if we do not admire it for its own sake. So, there is a reason for us to admire the evil demon for its own sake. But the evil demon is not good or of value.[6]

These examples are supposedly examples in which a saucer of mud and an evil demon satisfy the conditions on the right-hand side of the 'if and only if' in the bi-conditional *FA*, but are not good or of value.

The supposed reasons to admire the demon for its own sake and to desire the saucer of mud for its own sake are wrong kind of reasons because whatever these supposed facts are that are reasons to admire the demon and desire the saucer of mud they are not facts that make the demon or the saucer of mud good or of value—because the demon and the saucer are not good or of value. The wrong kind of reason (WKR) problem for FA accounts is the problem of restricting the bi-conditional *FA* so that it excludes the supposed reasons to admire the demon and desire the saucer of mud for their own sake—wrong kind of reasons—and so does not entail that the demon and the saucer are good or of value.

[1] See Rabinowicz and Rønnow-Rasmussen (2004, pp. 391–423), Lang (2008), Scanlon (1998, pp. 95–98), Stratton-Lake and Hooker (2006, pp. 152–153), and Way (2013).

[2] See Parfit (2011, pp. 38–39), McDowell (1985, p. 118), and Wiggins (1987, p. 206).

[3] See Brentano (1969, p. 18) Ewing (1947, chap. 5), Rabinowicz and Rønnow-Rasmussen (2004, pp. 394–400) and Suikkanen (2009, p. 768).

[4] See Stratton-Lake and Hooker (2006) and Way (2013).

[5] Hereafter I will use 'good' and 'of value' to refer to non-instrumental goodness *simpliciter* and final value. Something is non-instrumentally good *simpliciter* or of final value if it is of non-instrumental non-attributive value; see Korsgaard (1983).

[6] See Rabinowicz and Rønnow-Rasmussen (2004, pp. 405–407).

Many proponents of FA accounts have attempted to solve the WKR problem by restricting the scope of the bi-conditional *FA* in various ways. However, reasonable objections have been made to all extant attempts to solve the WKR problem in this way.[7] In this paper I argue that the WKR problem can instead be dissolved: I argue that there are no wrong kind of reasons such as reasons to admire the demon for its own sake or to desire the saucer of mud for its own sake.[8]

I argue that we cannot hold onto all three of the following plausible claims:

(A) There are normative reasons for us to admire a demon for its own sake and to desire a saucer of mud for its own sake when a demon will punish us if we do not.

(B) *Normative/Motivating Weak*. *R* is a normative reason for *A* to ϕ only if it is possible for someone to ϕ for the reason that *R*.

(C) *Generality*. *R* is a feature of *X* that is a reason for *A* to admire or desire *X* for its own sake only if, if *Y* has *R*, then other things being equal, *R* is a feature of *Y* that is a reason for *A* to admire or desire *Y* for its own sake (at least if *A* knows that *Y* has *R*).

And, I argue that we should hold onto (B) and (C) rather than (A).

There are two possibilities as to what the reason to admire the demon for its own sake in *Admire The Demon* and the reason to desire the saucer of mud for its own sake in *Desire a Saucer* are. Either (i) the reasons for us to admire the demon for its own sake and to desire the saucer of mud for its own sake are the facts that the demon will punish anyone who does not do these things, or (ii) these reasons are provided by features of the demon and the saucer of mud. So, if (A) holds, that is, if there are such reasons, then either (i) or (ii) holds. In Sect. 2 I argue that it is logically impossible for us to admire a demon for its own sake or desire a saucer of mud for its own sake for the reason that a demon will punish us if we do not. In Sect. 3 I argue that because of this logical impossibility (i) conflicts with (B) and that (B) is extremely plausible. In Sect. 4 I argue that (ii) conflicts with (C) and that (C) is extremely plausible.

In Sect. 5 I argue that we should give up (A) rather than giving up (B) or (C) because there is a good explanation of why (A) seems so plausible but does not in fact hold, but there is no similar good explanation of why (B) or (C) seem so plausible but do not in fact hold. The explanation of why (A) seems so plausible but does not in fact hold is that we mistake our intuitions about there being reasons to desire that we admire the demon for its own sake, reasons to bring it about that we admire the demon for its own sake, and reasons *why* it would be good to admire the

[7] See Rabinowicz and Rønnow-Rasmussen (2004, pp. 404–422, 2006, pp. 114–120), Stratton-Lake (2005), Lang (2008, pp. 475–484), Olson (2009, pp. 226–228), Schroeder (2010), Heuer (2011, pp. 169–173), Schroeder (2012, p. 465), and Way (2012).

[8] Skorupski (2007, pp. 9–12) and Parfit (2001, pp. 24–27, 2011, chap. 2 and Appendix 1) also hold that there are no such wrong kind of reasons; see also Rabinowicz and Rønnow-Rasmussen (2004, pp. 411–414). Jonathan Way (2012) has recently given a sustained argument for this conclusion. I take the argument in this paper to be a new argument for the view that Skorupski, Parfit, and Way hold.

demon for its own sake, for intuitions about there being reasons to admire the demon for its own sake. I explain just how easy a mistake this is to make.

2 Having a pro-attitude towards *X* for *X*'s own sake

Intuitively, if there is a reason for us to admire the demon for its own sake in *Admire the Demon* and a reason for us to desire a saucer of mud for its own sake in *Desire a Saucer of Mud*, the reasons for us to do these things are that the demon will punish us if we do not. In this section I will argue that

(I) It is logically impossible for us to admire a demon for its own sake or desire a saucer of mud for its own sake for the reason that a demon will punish us if we do not.

I propose that

CONSTRAINT. To the extent that *A* is having a pro-attitude towards *X* for *X*'s own sake *A* is not having a pro-attitude towards *X* for the additional consequences of having this pro-attitude towards *X*.

According to *CONSTRAINT*, *A* is having a pro-attitude towards *X* for *X*'s own sake only if *A* is *not only* having a pro-attitude towards *X* because of the additional consequences of their having this pro-attitude towards *X*.

There are several reasons to accept *CONSTRAINT*. Firstly, *CONSTRAINT* is extremely intuitively plausible. If I desire pleasure for its own sake, the reason for which I desire pleasure is not provided by the consequences of my desiring pleasure for its own sake but rather by the nature of pleasure. I desire pleasure for its own sake only if my reason for desiring pleasure is about the nature of pleasure, how it feels for instance. Similarly, to the extent that I desire friendship for its own sake the reason for which I desire it is not provided by the consequences of my desiring friendship, such as that this might lead me to have friends and be happy. If I desire friendship for its own sake, I see friendship itself as giving me a reason to want it. And I desire it for this reason. If you tell me that you hope that the rainforests are preserved for their own sake *just for the reason that* if you hope for this, then maybe other people will start hoping for the same thing, it seems that you've misspoken. Rather it seems that what you said was elliptical and you meant that you hope that the rainforests are preserved for their own sake *and hopefully* your hoping for this will lead others to hope for the same. It seems that you cannot hope for something for its own sake *just for the reason that* your hoping for that thing will produce good effects.

Similarly, suppose that I tell you that I admire John Rawls for his own sake on account of his manner and generosity, his creativity, ingenuity, and hard work but only because admiring him on the basis of these features will make me try to emulate Rawls and will increase the likelihood that I can be as successful and as esteemed as Rawls was. It seems that I've contradicted myself. Given my addition that I admire Rawls *only because* doing this will make me try to emulate him and

increase the likelihood that I will be successful it seems that I don't in fact admire him for his own sake. I may perhaps admire him but I don't admire him for his own sake.

A second reason to accept *CONSTRAINT* is that it flows from the most plausible account of what it is to do something for its own sake. It seems that

> *REDUCTION*. Claims about whether *A* ϕd for its own sake or not are just, in addition to claims about *A*'s ϕ-ing, claims about the reasons for which *A* ϕd.

When we contrast someone who keeps a promise for its own sake with someone who does not keep a promise for its own sake it seems that we are contrasting the reasons for which these agents did what they did; we are contrasting the kind of reasons for which the person who kept a promise for its own sake kept her promise with other kinds of reasons for which she could have kept it such as to avoid disappointing others, to avoid losing the trust of others, or to avoid punishment. When we talk about people wanting or pursuing friendship, democracy, power, or fame for its own sake we are meaning to contrast the reasons for which these people are pursuing these things with other reasons for which they might have been pursuing these things—namely as a means to something else such as happiness or influence.[9]

Given *REDUCTION*, we should distinguish four types of for its own sake/not-for its own sake claims and four types of reason for which *A* might have had a pro-attitude that these different claims pick out:

(a) *A* might have ϕd for *X*'s sake—as when I admire someone because they are kind and generous. In this case *A*'s reason for ϕ-ing was a feature of *X*;
(b) *A* might have ϕd for the sake of of *X*'s consequences—as when I desire a plane ticket because I desire to lie on a beach for a week. In this case *A*'s reason for ϕ-ing was (a feature of) the consequences of (getting) *X*;
(c) *A* might have ϕd for the sake of ϕ-ing—as when I keep a promise for the sake of keeping this promise. In this case *A*'s reason for ϕ-ing was ϕ-ing itself, the nature of ϕ-ing or a feature of ϕ-ing itself; or
(d) *A* might have ϕd for the sake of the consequences of ϕ-ing—as someone who keeps a promise for the consequences of keeping this promise does and as someone who approves of a demon for the reason that if they do not approve of the demon, then the demon will punish them does, supposing that this is possible. In this case *A*'s reason for ϕ-ing was (a feature of) the consequences of ϕ-ing.

Although it is possible for *A* to do two or more of (a–d) at the same time it is not possible for *A* to do one to the extent that she does another. That is, it is not possible that to the extent that *A*'s reason for ϕ-ing is a feature of *X*, *A*'s reason for ϕ-ing is also a feature of the consequences of ϕ-ing. For instance, if *A*'s only reason for ϕ-ing is a feature of the consequences of her ϕ-ing, then *A*'s only reason for ϕ-ing is

[9] Of course, it might be that the people whom we are talking about pursue friendship or power for several reasons in which case we are contrasting one type of reason for which they are pursuing friendship or power with another type of reason for which they are also pursuing friendship or power.

not a feature of *X*. It's not possible to *ϕ for only one reason* but at the same time to *ϕ for more than one reason*.

To be clear, it does not follow from *REDUCTION* that it is not possible for someone to do both (a) and (d) at the same time. This might be possible. It might be possible to simultaneously desire a saucer of mud for its own sake—because of features of the saucer of mud such as its grittiness and beauty [sic]—and to desire a saucer of mud (not for its own sake) because a demon has threatened to punish you if you do not. *REDUCTION* is agnostic on this issue. All *REDUCTION* entails is that it is not possible to desire a saucer of mud in virtue of features of the saucer of mud, such as its grittiness and beauty [sic], for the reason that a demon has threatened to punish you if you do not. This would not merely be to do (a) and (d) at the same time but to do (a) to the extent that one is doing (d) and to do (d) to the extent that one is doing (a). And if we should understand claims about *A*'s *ϕ*-ing for its own sake in terms of the reasons for which *A ϕ*s, then this is logically impossible: there could only be two things going on here not one.

Seeing this conclusion many will want to deny *REDUCTION*. But there are problems with denying *REDUCTION*. The view that '*A ϕ*s for its own sake' says something more about *A's ϕ*-ing than just something about the reasons for which *A ϕ*s is unmotivated. Furthermore, the view that '*A ϕ*s for its own sake' says something more about *A's ϕ*-ing than just something about the reasons for which *A ϕ*s fails to explain why we frequently alternate between talking about an agent's doing something for its own sake and specifying the reasons for which that agent did that thing. For instance, we alternate between talking about someone keeping a promise for its own sake and talking about the reasons for which they kept their promise, which are not reasons associated with the good additional consequences of so keeping their promise. In contrast, *REDUCTION* explains this strong connection between talk about whether *A ϕ*d for its own sake or not and talk about the reasons for which *A ϕ*d by reducing talk about whether *A ϕ*d for its own sake or not to talk about the reasons for which *A ϕ*d. And since strong connections call for explanation, this is a clear advantage of *REDUCTION* and a clear disadvantage of competing views, which cannot explain this strong connection.[10]

So there are several reasons to accept

> *CONSTRAINT*. To the extent that *A* is having a pro-attitude towards *X* for *X*'s own sake *A* is not having a pro-attitude towards *X* for the additional consequences of having this pro-attitude towards *X*.

And *CONSTRAINT* (trivially) entails that

(I) It is logically impossible for us to admire a demon for its own sake or desire a saucer of mud for its own sake for the reason that a demon will punish us if we do not.

However, *CONSTRAINT* and (I) may *appear* counterintuitive. Firstly, it seems that we can desire things for their own sake *because* of the additional consequences

[10] On why strong connections call for explanation, see, for instance, Enoch (2011, p. 158).

of doing so. For instance, it seems that we can a desire a saucer of mud *because* a demon will punish us if we do not. But *CONSTRAINT* and (I) seem to entail that we cannot desire a saucer of mud *because* a demon will punish us if we do not. However, this is not the case. It is consistent with *CONSTRAINT* and (I) that in one clear sense we can desire a saucer of mud *because* the demon will punish us if we do not.

Suppose that we are aware that the demon will punish us if we do not desire a saucer of mud for its own sake. We can try to make ourselves desire a saucer of mud for its own sake. And if we succeed in desiring a saucer of mud for its own sake—i.e. by desiring a saucer of mud for its aesthetic qualities—then this will be *because* we tried and succeeded in making ourselves desire the saucer of mud for its own sake, by, for instance, getting ourselves hypnotized. And our reason for trying to desire the saucer of mud for its own sake will have been that the demon would punish us if we did not desire the saucer of mud for its own sake.

So we can *in a sense* desire a saucer of mud for its own sake *because* the demon would punish us if we did not without desiring a saucer of mud for its own sake for the reason that a demon would punish us if we did not. But only in the same way that we can believe in vegetarianism *because* of an experience that we had at a factory farm without believing in vegetarianism *for this reason*. And only in the same way that we can be (particularly) annoyed *because* we have not had a coffee yet this morning without the fact that we have not had a coffee yet being *the reason for which* we are annoyed (we might be annoyed that the mail hasn't come or by what is being said on the radio for instance).

In all these cases the sense of 'because' is that of an explanatory reason, that which explains why we are doing something in a way that does not make reference to and differs from, the reason for which we do something. (For instance, bridges can fall for reasons and because of things in these senses of 'reason' and 'because' but bridges don't have motivating reasons to fall, that, is they do not have reasons for which they fall—they do not act *for* reasons).[11] When I use this sense of because to say that I believe in strongly redistributive taxation because of my left-wing upbringing I don't mean that I see my left-wing upbringing as a reason to believe in strongly redistributive taxation or that this is my ground for believing in redistributive taxation. Rather I mean that my left-wing upbringing is what explains why I believe in redistributive taxation for the reasons that I do. So, it is not incompatible with *CONSTRAINT* and (I) that we can desire a saucer of mud for its own sake *because* (in the sense of, what explains why we are doing this rather than the reason for which we are doing this) the demon has threatened to punish us if we do not. And accepting that we can desire a saucer of mud for its own sake *because* of a demon's threat in this sense is not accepting that we can desire a saucer of mud for its own sake *for* the reason that a demon has threatened us.

Secondly, suppose that an evil demon tells us:

> Unless you desire this saucer of mud for its own sake, I will torment you forever. But you must not, for instance, get yourself into a state in which you

[11] Cf. Dancy (2000, pp. 6–7).

> desire the saucer of mud on the basis of features of the saucer such as its aesthetic qualities, and thereby desire the saucer of mud for its own sake. Rather you must desire the saucer of mud for its own sake *only* for the reason that I will torment you forever if you do not.

According to *CONSTRAINT* and (I) the demon is demanding the logically impossible since it is logically impossible to desire the saucer of mud for its own sake *only* for the reason that it will torment you forever. It seems intuitive to me that what the demon demands here is the logically impossible. But perhaps it does not seem intuitive to others that the demon is demanding the impossible in this case. These people will hold that *CONSTRAINT* and (I) are counter-intuitive because they entail that in this case the demon is demanding the impossible.

However, it might be that the demon is demanding the impossible even though it does not seem to some that he is. One reason that it might seem counter-intuitive that the demon is demanding the impossible even though he is in fact demanding the impossible is that there is pressure to charitably interpret the demon. There's pressure to interpret the demon as demanding something that we at least logically can do, for why would he demand something logically impossible when we cannot even begin to try to do something logically impossible?

Furthermore, it may be that what the demon is demanding is logically impossible but un-obviously logically impossible. If a demon demands that we make a round square, it is obvious that the demon is demanding the impossible and so it is intuitive that the demon is demanding the impossible. But other things are not so obviously logically impossible. Suppose that the demon demands that we believe that it is raining and believe that there is no reason or evidence that it is raining at the same time. It is not obvious that the demon is demanding the logically impossible in this case, and so we do not intuitively believe that the demon is demanding the impossible, but nevertheless there are good reasons to believe that the demon is in fact demanding the logically impossible.[12] Or suppose that the demon demands that we pair *all* the natural numbers with *all* the real numbers between 0 and 1. It is not obvious that the demon is demanding the logically impossible, and so it is not intuitive that the demon is demanding the logically impossible. But in fact he is demanding the logically impossible.[13] It may be that some people's intuitions about the impossibility of the demon's demand that we desire a saucer of mud for its own sake *only for* the reason that he will punish us if we do not are off-track in the same way that our intuitions about the impossibility of the demon's demands that we do these other things are off-track. That is, it may just be that the demon is demanding the un-obviously impossible.

But even if the explanations that I have provided do not sufficiently explain why it seems counter-intuitive to some people that the demon is demanding the impossible in this case, in order to hold onto the view that their intuitions are tracking the truth those who find it counter-intuitive to hold that the demon is demanding the impossible in this case need to discharge several explanatory debts.

[12] See, for instance, Streumer (2013, pp. 196–199).

[13] Presuming that the relevant parts of Cantor's diagonalization argument are sound.

They need to: (i) provide an account of what it is for *A* to ϕ for its own sake other than in terms of reasons; (ii) explain the strong connection between talk about agents ϕ-ing and not ϕ-ing for its own sake and talk about the kinds of reasons for which agents ϕ and do not ϕ; and (iii) explain why it seems contradictory for me to claim that, for instance, I admire John Rawls for his own sake on account of his manner and generosity but only for the reason that admiring him on the basis of these features will increase the likelihood of me being as successful as Rawls was. Until these debts have been discharged the burden of proof lies with those who hold that the demon is not demanding the impossible and who deny *CONSTRAINT* and *REDUCTION* to show that the demon is not demanding the impossible.

3 The demon's threat and the relationship between normative and motivating reasons

In the last section I established that

(I) It is logically impossible for us to admire a demon for its own sake or desire a saucer of mud for its own sake for the reason that a demon will punish us if we do not.

Now, I will show that since (I) holds, the idea that the reason for us to admire the demon for its own sake when it will punish us if we do not is the fact that it will punish us if we do not conflicts with an extremely weak but extremely plausible claim about the relationship between normative and motivating reasons.

There are several types of reasons. FA Accounts, *Desire a Saucer of Mud*, and *Admire the Demon* deal with *normative reasons*. A normative reason to ϕ is a consideration that counts in favour of ϕ-ing: that there will be dancing at a party is a normative reason for me to go to the party if I enjoy dancing. But that there will be dancing at the party might not be the reason *for which* I go to the party—it might not be my *motivating reason* to go to the party; I might go to the party unaware that there will be dancing because I promised a friend that I would go or because a woman I like is going to be there.

Most philosophers hold that there is a relationship between normative and motivating reasons. According to Bernard Williams and John Skorupski

Normative/Motivating Strong. *R* is a normative reason for *A* to ϕ, only if it is (in some sense) possible for *A* to ϕ for the reason that *R*.[14]

And according to Joseph Raz and Jonathan Dancy,

(B) *Normative/Motivating Weak*. *R* is a normative reason for *A* to ϕ only if it is possible for *someone* to ϕ for the reason that *R*.[15]

[14] See Williams (1989, pp. 38–39) and Skorupski (2010, pp. 73–74).

[15] See Raz (2011, p. 27) and Dancy (2000, p. 101).

(Neither of these claims entails Williams's internalism about reasons, although *Normative/Motivating Strong* is often used as a premise in arguments for Williams-style internalism about reasons. It is consistent with *Normative/Motivating Weak*, for instance, that there are normative reasons for *A* to do things that she could not be motivated to do.)

One reason to hold some version of *Normative/Motivating Strong* is that it explains why there are no reasons for non-rational animals. The reason why there are no reasons for these animals is that they cannot respond to reasons and so cannot perform actions for reasons; the claim that *R* is a reason for *A* to ϕ only if *A* can ϕ might not explain why there are no reasons for non-rational animals since non-rational animals can perhaps perform actions.

Normative/Motivating Strong entails *Normative/Motivating Weak* And there are several reasons why we should hold at least *Normative/Motivating Weak*. Firstly, *Normative/Motivating Weak* preserves the intuition that normative reasons must be capable of guiding us; a normative reason to ϕ that could not possibly be anyone's reason for ϕ-ing could not play this guiding role.[16] *Normative/Motivating Weak* also preserves the idea that there is a tie between reasons and advice. When we tell a friend that they should pay back a loan because if they don't, they won't be able to get credit in the future we think that the reason we give them serves as advice. But this reason's status as advice would be undermined if our friend were incapable of paying back the loan for this reason.[17] Furthermore, if even the weak relationship between normative and motivating reasons specified by *Normative/Motivating Weak* did not hold, then it would just be an accident of etymology that normative reasons are called reasons and motivating reasons are also called reasons; these types of reasons would have as little in common as 'bank' does in financial bank and riverbank.[18] But normative and motivating reasons seem to be more closely related than financial banks and riverbanks are.[19]

But if *Normative/Motivating Weak* holds, then the fact that a demon will punish us if we do not admire him for his own sake could not be a reason for us to admire him for his own sake. As I argued in Sect. 2, it is logically impossible for us to admire a demon for its own sake or desire a saucer of mud for its own sake for the reason that a demon will punish us if we do not. And according to *Normative/Motivating Weak*, if it is not logically possible for anyone to ϕ for the reason that *R*, then *R* is not a normative reason to ϕ. So, it seems that I have established that

[16] See Markovitz (2011, p. 149).

[17] See Shah (2006, p. 486).

[18] See Markovitz (2011, p. 148).

[19] According to Markovitz (2011, p. 153) (the contents of) our current unjustified false beliefs are reasons for us to believe that we are fallible. But we could not believe that we are fallible on the grounds of the (contents of) these beliefs because to do so would be to no longer have those beliefs. However, *others can* believe that we are fallible on the grounds of our beliefs, for they are not the one's who have our beliefs. So, even if Markovitz's example is a counter-example to *Normative/Motivating Strong* it is not a counter-example to *Normative/Motivating Weak*.

(II) The intuitive idea that the reason for us to admire the demon for its own sake in *Admire the Demon* and desire a saucer of mud for its own sake in *Desire a Saucer of Mud* is the fact that the demon will punish us if we do not do these things conflicts with a deeply plausible and weak thesis about the relationship between normative and motivating reasons, namely (B) *Normative/Motivating Weak*.

In response to my argument for (II) it might be argued that the reason for us to admire the demon for its own sake is not that it will punish us if we do not but rather that the demon has *threatened* us. And admiring a demon because it has threatened you is not admiring a demon for the additional consequences of admiring it.[20] So, the reason for us to admire the demon does not conflict with (B), *Normative/Motivating Weak*, because the reason for us to admire it for its own sake is that it has threatened us and it is possible to admire the demon for its own sake because it has threatened us.

But if the demon's threat was an empty threat and we were aware of this, then that it has threatened to punish us if we do not admire it for its own sake would be no reason for us to admire it. So even if the fact that the demon has threatened to punish us if we do not admire it for its own sake is part of the reason for us to admire it for its own sake, it cannot be the whole reason. Rather the complete reason for us to admire it for its own sake would have to be that the demon has threatened us and there is some non-zero probability that it will punish us if we do not admire it for its own sake. But if this is the reason for us to admire the demon for its own sake, then it is not a reason for which we could admire the demon for its own sake, since it essentially refers to the consequences of our admiring the demon for its own sake.[21] And, as I argued, to the extent that we are admiring something for its own sake we are not admiring it even partially because of the consequences of our admiring that thing; it is logically impossible to admire something for its own sake for the reason that doing this will lead to good consequences. So, this reason also conflicts with (B) *Normative/Motivating Weak*. And so, this response does not undermine my argument for (II).

4 Features of the demon and the saucer and the generality of reasons to desire and admire

It might be that the reason for us to admire the demon for its own sake when the demon will punish us if we do not is *not* that the demon will punish us if we do not admire it for its own sake. Instead it might be that the fact that the demon will punish us if we do not admire it for its own sake makes other facts, such as the demon's power, into reasons to admire the demon. Similarly, it might be that the fact that the demon will punish us if we do not desire a saucer of mud for its own

[20] See Rabinowicz and Rønnow-Rasmussen (2004, pp. 419–420).

[21] Alternatively, it might be that to do something because something has threatened to punish you if you do not *just is* to do that thing for the additional consequences of doing that thing.

sake is not a reason for us to desire a saucer of mud for its own sake but rather makes other facts, such as its grittiness, its texture, or its beauty [sic], into reasons to desire a saucer of mud for its own sake. On this view the demon's threat *generates* reasons for us to desire a saucer, or *confers reason-providing status* on other features of a saucer of mud.[22] Call the view that the demon's threat makes *other features* of the saucer of mud into reasons to desire it for its own sake the *reason-conferring* view of the reason to desire the saucer of mud for its own sake. But as I'll show, the reason-conferring view of the reason to desire the saucer of mud for its own sake conflicts with the generality of reasons to admire things for their own sake and reasons to desire things for their own sake.

Reasons to admire or desire something for its own sake have a certain *generality*, namely:

(C) *Generality*. *R* is a feature of *X* that is a reason for *A* to admire or desire *X* for its own sake only if, if *Y* has *R*, then other things being equal, *R* is a feature of *Y* that is a reason for *A* to admire or desire *Y* for its own sake (at least if *A* knows that *Y* has *R*).[23]

If features or traits of an athlete or virtuous person are reasons for us to admire them (for their own sake), then, other things being equal, there is a reason for us to admire other athletes or virtuous people who have these features or traits. If someone's having performed an heroic act is a reason to admire them, then other things being equal there is a reason to admire someone who has performed the same heroic act. There might be two tennis players who have equal tennis playing skills, but there might be a reason for us to admire one and not the other. If this is so, there must be some feature of the player whom there is no reason for us to admire that makes other things not equal. Perhaps there is a reason for us to admire Federer as a tennis player, but no reason for us to admire McEnroe even though we take them to have the same tennis playing skills. But, then there is no reason for us to admire McEnroe because other things are not equal between McEnroe and Federer; McEnroe's bad sportsmanship makes other things not equal between him and Federer and prevents there from being a reason for us to admire him.

Similarly, suppose the fact that a painting has a certain texture, certain colours, or a certain holistic unity is a reason for me to desire it for its own sake. If, perhaps impossibly, I found a painting with exactly the same features, there would be a reason for me to desire it for its own sake. Or suppose that there is a reason for me to desire that some serial killer is punished because of what he did (for its own sake). If I hear of another serial killer who did the same things, then, other things being

[22] We might say that the demon's threat is 'backgrounded' on this view; see Schroeder (2007, chap. 2).

[23] I add the caveat, at least if *A* knows that *Y* has *R*, because it might be that there are no reasons for us to admire people that have the same characteristics as people that we have reason to admire but whom, unlike the people whom we have reasons to admire, we are completely unaware of. Imagine that Bryony bravely jumps onto the subway tracks and saves you from the oncoming train. It may be that you have reason to admire Bryony but you do not have reason to admire Becky who bravely jumped onto the subway tracks and saved someone else's life because you are not aware that Becky did this. See, for instance, Dancy (2000, pp. 56–58). *Generality* is neutral on this issue.

equal, there's a reason for me to desire (for its own sake) that he is punished because of what he did too.

To be clear, *Generality* is not in conflict with the view that some facts confer reason-providing status on features of things (or are background conditions on reasons). So it is consistent with *Generality* that desires might make other considerations into reasons. It is consistent with *Generality*, for instance, that the fact that there will be dancing at a party is a reason for Ronnie to go to that party, but not for Bradley because Ronnie likes dancing and Bradley does not, and that the fact that Ronnie likes dancing is not itself part of the reason for Ronnie to go to the party—but only makes the fact that there will be dancing at the party into a reason for Ronnie to go. This is because *Generality* is a claim about the reasons that there are for particular agents not a claim about the reasons that there are for one agent given the reasons that there are for another agent; we might say that it's a claim about intrapersonal reasons rather than about interpersonal reasons. All that follows from *Generality* with regards to Ronnie and Bradley, for instance, is that the fact that there is dancing at the party is a reason for Ronnie to desire to go to the party only if, if there is dancing at another party, then other things being equal, the fact that there is dancing at this party is a reason for Ronnie to go to this other party.

So, (C) *Generality* seems intuitively plausible. But if (C) *Generality* holds, the reason-conferring view of the reason to admire the demon—that is, that the demon's threat makes other features of the demon, such as his power, into reasons to admire him for his own sake—leads to extremely counter-intuitive consequences.

Consider the following case:

> *Evil Demon and Evil Cat*. An evil demon will severely punish us if we do not admire it. *A* is one of us. But there is an evil cat in another world, which is a world that *A* has knowledge of, but which *A* cannot communicate with, and which is identical to *A*'s world except that *A* is not in that world—although her counterpart is—and that instead of an evil demon there is an evil cat, and the evil cat will punish people in that world who do not admire it.

Even if *A* knows about the evil cat, we do not think that she has a reason to admire the evil cat when the evil demon has threatened her with punishment if she does not admire it.

But if the demon's threat makes features that would not otherwise be reasons to admire the demon into reasons to admire it, then

(1) Whatever property the evil demon has that provides a reason for *A* to admire it, the evil cat has that property too.[24]

But (2) follows from (1) and (C) *Generality*:

[24] If you don't accept this because the cat is a cat, we can modify the example so that the world is a counterpart world in which everything is identical to this world, so that there is no evil cat, but just another evil demon. If you don't accept this because the demon is in another world, then modify the example to one in which the demon is in half of our divided world threatening us in that half of our world and another evil cat/demon is in the other half of our divided world threatening the other half of the world.

(2) There is a reason for *A* to admire the evil demon only if there is a reason for *A* to admire the evil cat.

So, the combination of (C) and the idea that the demon's threat makes features that would not otherwise be reasons to admire the demon into reasons to admire it generates the untenable result that there is a reason for *A* to admire the evil cat.[25]

It might be objected that (2) does not follow from (1) and (C) because other things are not equal between features of the evil demon, which is in this world, and features of the evil cat, which is in the other world, because there is a reason for *A* to admire the demon because the demon will punish *A* if she does not, but the evil cat will punish *others*—not *A*—in the other world if they do not admire it. But although your admiring something can causally depend on that thing's relation to you, the reason for you to admire something cannot depend on that thing being related to you. The reason for you to admire something cannot depend for its status as a reason for you to admire it on that thing's relation to you.[26]

Consider the following example:

Subway Hero. You collapse at a subway station overtaken by convulsions. You manage to raise yourself but then stumble and fall onto the subway tracks. A 50 year old construction worker leaps onto the tracks and presses his body down on top of you pushing you into a trough about a foot deep. The subway train attempts to stop. Five cars screech over the top of you and the construction worker before the train stops. You both survive, the construction worker having pressed you deep enough into the trough for you both to be saved. When you return to a state of normality there is a reason for you admire the construction worker for his bravery. You then learn of another incident on the other side of the world in which someone else stumbled onto train tracks and was pushed into a trough and saved by a construction worker. All you know about the two construction workers are the facts stated.[27]

To the best of your knowledge the other construction worker has the same features in virtue of which there is a reason for you to admire the construction worker who saved you. And it seems that since there is a reason for you to admire the construction worker who saved you, there is a reason for you to admire the

[25] It might be claimed that appealing to *Generality* is begging the question against the idea that the demon's threat confers reason-providing status on features of the saucer of mud given the obviously counter-intuitive consequences of the combination of *Generality* and the reason conferring view of the demon's threat. However, I have argued that *Generality* seems plausible in all other cases, so it is ad-hoc to claim that it does not hold in the case of the demon's threat, at least not without providing an explanation of why it does not hold in the case of the demon's threat.

[26] And *A*'s relationship with the demon could not figure in the content of the reason for her to admire the demon consistent with the reason-conferring view of what the reason for *A* to admire the demon is. If her relationship to the demon figured in the content of the reason for her to admire the demon, the reason for her to admire the demon for its own sake would be that he will punish her if she does not, and as I argued in Sect. 3, the idea that this is the reason for *A* to admire the demon for its own sake conflicts with (B) *Normative/Motivating Weak*.

[27] This is modelled on the heroism of Wesley Autrey. See Buckley (2007).

construction worker on the other side of the world even though the construction worker on the other side of the world *did not save you.*

Furthermore, suppose there is a reason for you to admire your daughter because of her work-ethic. Your friend's daughter has a similar work-ethic. You know this. If all other things are equal, and suppose they are, there is a reason for you to admire your friend's daughter. There might be a reason for you to be proud of your daughter but no reason for you to be proud of your friend's daughter. However, there could not be a reason for you to admire your daughter if there was no reason for you to admire your friend's daughter.

It might seem that there are examples in which a reason for you to admire someone can depend on their relationship to you. For instance, if you cheat a stranger and a friend and both forgive you, it seems that you would reasonably admire your friend more than the stranger. But in this case the strength of the reason for you to admire your friend for forgiving you depends on her *forgiving a friend* and not on her *forgiving you*. There is a reason for you to admire your friend more than the stranger because she has forgiven a friend and there would be a reason for you to admire her more than the stranger whether or not that friend happened to be you.[28]

Similarly, the combination of the reason-conferring view of the reason to desire a saucer of mud for its own sake when a demon will punish you if you do not and (C), *Generality*, leads to counter-intuitive consequences.

Desire a Saucer of Mud is always elaborated as a case in which there is a reason for us to desire *a* saucer of mud for its own sake. But this is ambiguous between a reason to desire *any old* saucer of mud, or a saucer of mud *in the abstract*, and a reason to desire *a particular* saucer of mud. It's never clarified whether the demon will punish us unless we desire *a particular* saucer of mud for its own sake, or a saucer of mud *in general* because this really doesn't seem to matter. If we suppose that the demon will punish us if we don't desire *a particular* saucer of mud, then it seems that there is a reason for us to desire a particular saucer of mud. If we suppose that the demon will punish us if we don't desire a saucer of mud *in general*, then it seems that there is a reason for us to desire a saucer of mud *in general*.

Now suppose that the demon will punish us if we don't desire a particular saucer of mud, saucer of mud 1, for its own sake. *A* is one of us. But *A* knows about another

[28] In response it might be argued that we can provide a new evil demon case that does not conflict with (C). We can imagine a case in which an evil demon will punish *A* unless she admires it and every demon, cat, or whatever, that is just as evil and powerful as the demon. And it is not counter-intuitive to claim that there is a reason for *A* to admire the evil cat for its own sake in this new case. However, my argument was never intended to show that we cannot imagine a case that seems like a wrong kind of reason case. Rather my argument provides us with grounds to doubt our intuitions in the original case. That we can imagine a case in which a demon, understanding the generality constraint on reasons to admire and desire things for its own sake, makes a threat that is specified in such a way to satisfy this constraint, does not undermine my argument. This is because, if we have grounds to doubt our intuitions about the original case in which it seemed that there was a reason to admire the demon for his own sake when he would punish us if we did not, then we should also doubt our intuitions in similar cases including a case in which the demon will punish us unless we admire him *and* do other things too. If, *contra* our intuitions, in all the cases in which it seemed that there was a reason to admire a demon due to the demon's threat there was no such reason, then the view that there is a reason to admire the demon in such a new case because the demon's threat is backgrounded would need to be motivated.

saucer of mud, saucer of mud 2, which is identical to saucer of mud 1—in texture, grittiness, and all other ways. We don't think that there's a reason for *A* to desire saucer of mud 2 for its own sake. After all, the demon will only punish her if she does not desire saucer of mud 1 for its own sake, not if she does not desire saucer of mud 2 for its own sake.

But if the demon's threat makes features that would not otherwise be reasons to desire the saucer of mud for its own sake into reasons to desire it for its own sake, then

(1*) Whatever property saucer of mud 1 has that provides a reason for *A* to desire it saucer of mud 2 has that property too.

And (2*) follows from the combination of (1*) and (C), *Generality*:

(2*) There is a reason for *A* to desire saucer of mud 1 for its own sake only if there is a reason for *A* to desire saucer of mud 2 for its own sake.

It might be objected that (2*) does not follow from (1*) and (C) because other things are not equal between features of saucer of mud 1 and saucer of mud 2 since the demon will punish *A* if she does not desire saucer of mud 1 for its own sake but will not punish *A* if she does not desire saucer of mud 2 for its own sake. But as with admiring someone or something for their/its own sake, the reason for you to desire something for its own sake cannot depend for its status, as a reason for you to desire that thing for its own sake, on that thing's relationship to you. Rather, if *X* and *Y* are identical except that you bear a particular relationship to *X*, and there is a reason for you to desire *X* for its own sake but not *Y*, then your relationship to *X* figures in the reason for you to desire *X* for its own sake.

Suppose that there are two serial killers who have committed the same crimes and there is a reason for you to desire the punishment of one of the serial killers for its own sake, but no reason for you to desire the punishment of the other serial killer for its own sake. And suppose that this difference is due to the fact that one of the serial killers killed your relatives and the other did not. If this were the case, then it would not be that the features of the two serial killers were the same but other things were not equal with regards to them. Rather, the reason for you to desire that one of them is punished for its own sake would be that he killed *your relatives*, and the other serial killer would not have the property of having killed *your relatives*. So, one of the killers would have a property that provided a reason for you to desire that he is punished for its own sake, which the other killer did not have. And this property would give the reason for you to desire that he is punished for his own sake; this property would not merely be that factor in virtue of which other things are not equal.

Similarly, there might be two pieces of countryside that are exactly the same but there might be a reason for me to desire that one is preserved for its own sake and no such reason for me to desire that the other is preserved for its own sake. If this asymmetry is down to the fact that my ancestors are from one of these stretches of countryside, then that stretch of countryside has a property the other does not, namely it has the property of being where *my ancestors* are from. And this property of being where *my ancestors* are from is the property that provides me with a reason to desire that it is preserved for its own sake.

It might be claimed that *A*'s relationship to saucer of mud 1, namely that she will be punished if she does not desire the saucer of mud for its own sake is part of the reason for her to desire the saucer of mud for its own sake. But, if this is so, then the idea that there is a reason to desire saucer of mud 1 for its own sake conflicts with *Normative/Motivating Weak*. As I argued in Sects. 2 and 3, if part of the reason for *A* to desire saucer of mud 1 for its own sake is the fact that she will be punished if she does not, then if *A*, or anyone, desired the saucer of mud for this reason, she would not be desiring the saucer of mud for its own sake. And if *A* or anyone else could not desire the saucer of mud for its own sake for this reason, then the view that there is such a reason is in conflict with *Normative/Motivating Weak*.[29]

5 Explaining the error

I've been arguing that the following three plausible claims cannot all be true:

(A) There are normative reasons for us to admire a demon for its own sake and to desire a saucer of mud for its own sake when a demon will punish us if we do not.
(B) *Normative/Motivating Weak. R* is a normative reason for *A* to ϕ only if it is possible for someone to ϕ for the reason that *R*.
(C) *Generality*. *R* is a feature of *X* that is a reason for *A* to admire or desire *X* for its own sake only if, if *Y* has *R*, then other things being equal, *R* is a feature of *Y* that is a reason for *A* to admire or desire *Y* for its own sake (at least if *A* knows that *Y* has *R*).

In Sects. 2 and 3 I argued that one of the two ways in which (A) could be true conflicts with (B) and in Sect. 4 I argued that the other way in which (A) might be true conflicts with (C). I'll now argue that we should give up (A) rather than (B) or (C).

There is no good explanation of why (B) seems so plausible but is false. And as I argued in Sect. 3, there are several reasons to accept (B). Nor is there a good explanation of why (C) is so plausible, as I argued that it is in Sect. 4, but in fact false. And it would be ad-hoc to simply hold that (B) and (C) apply to all reasons to admire and desire things for their own sake apart from those in evil demon cases. In contrast, there's a very good explanation of why although we believe that there are reasons to admire the demon for its own sake and to desire the saucer of mud for its own sake there are no such reasons. Namely, that we mistake the intuition that there are normative reasons for us to *desire that we admire* the demon for his own sake, normative reasons for us to *try to bring it about that we admire* him for his own sake, or reasons *why* it would be *good for us to admire* the demon for his own sake

[29] It might be objected that we should not think of the reason to desire the saucer of mud for its own sake as a reason to desire only a particular saucer of mud for its own sake. But it didn't seem to matter whether the example was about a *particular* saucer of mud or a saucer of mud in *general*. So it's odd that it now does matter. See also *supra* note 28.

for the intuition that there are normative reasons for us *to admire* the demon for his own sake when he will punish us if we do not.[30]

As I see it, there are three factors that combine to make people believe that their intuitions are about reasons for us to admire the demon. First, the reason for us to desire that we admire the demon and the reason for us to bring it about that we admire the demon are extremely weighty. And when the reasons for us to *desire that we* ϕ and to *bring it about that we* ϕ are very weighty, and when it would be extremely good for us if we ϕ'd, we sometimes suppose that there must be a reason for us to ϕ when there could not be.

Suppose that an evil demon will punish us if we do not spontaneously turn ourselves into a biologically accurate killer whale. We cannot spontaneously turn ourselves into a biologically accurate killer whale, so there is no reason for us to spontaneously turn ourselves into one. However, many people would be unwilling to accept that there is no reason for us to spontaneously turn ourselves into a biologically accurate killer whale; and would reject the claim that it would *only* be good for us to be able to do this and that there is *only* a reason for us to wish that we could do this. It might be that 'there is no reason to be a biologically accurate killer whale' conversationally implies, 'there is no reason to desire that you are a biologically accurate killer whale' or 'there is no reason to wish that you were a biologically accurate killer whale.' But it is obviously the case that one can wish that one could do things that one cannot do, or were things that one is not. I can wish that I were a different ethnicity, lived in a different era, or could fly like a bird, for instance.

Secondly, there is no consensus as to whether it is possible to admire or desire something directly just by deciding to admire or desire that thing. This contrasts with other things that no one believes it is possible to do directly by deciding to do them. Consider the title of a current play, *Reasons to be Pretty*. There could not be a reason for anyone to *be* pretty, since we cannot be pretty directly just by deciding to be pretty—without going through some further process. So a reason for someone to *be* pretty would be like a reason for them to be a little bit taller, to have a headache, to be Michael Jordan, or to be a bat. We should hold that there are strictly no reasons to be pretty, and 'a reason to be pretty' is normally elliptical for 'a reason to make oneself pretty,' 'a reason to want to be pretty,' or 'a reason *why it would be good* to be pretty'—the title of the play seems to be a play on the latter.[31] But, unlike being pretty, there is no consensus as to whether we can directly admire or desire something for its own sake. There is no consensus as to whether we can simply admire something that we decide to admire directly just by deciding to admire it rather than having to go through some process, such as hypnosis or brainwashing, in

[30] See Skorupski (2007, pp. 9–12), Parfit (2001, pp 24–27, 2011, chap. 2 and Appendix 1), and Way (2012).

[31] Thus, although Daniellson and Olson (2007, pp. 513–514) might be right that it is sometimes odd to claim that there is a reason to try to ϕ when one cannot ϕ sometimes there certainly are reasons to try to ϕ when one cannot; such as to try to be pretty even though there is no reason to just spontaneously be pretty. And there are certainly reasons to wish that we could ϕ even if we cannot ϕ such as for us to wish that we could fly even though we cannot.

order to get ourselves to admire that thing. This lack of consensus as to whether it is possible to admire something perhaps throws our intuitions off-track.

Finally, there are several different senses of 'reason' that are somewhat similar—and which sometimes refer to the same things—that are easily confused, and that can confuse our intuitions. The reason that a bridge collapsed, that it was under too much weight, is, for instance, an explanatory reason; there was no normative reason *for* the bridge to collapse, but only a reason *why* the bridge collapsed.[32] Some people confuse explanatory and normative reasons in the cases that I've been discussing. Sometimes people who claim that there is a reason for us to be pretty, or to *be* a biologically accurate killer whale when a demon will punish us if we are not, are making claims about explanatory reasons: there is a reason *why* our being a biologically accurate killer whale would be good, and there might be reasons *why* being pretty would be good, but there are no (non-elliptical) reasons *for* being pretty or being a biologically accurate killer whale. So, it would not be odd if we did the same when it came to our intuitions about reasons in evil demon cases. These three factors lead us to mistake intuitions about reasons for us to desire that we admire the demon or to bring it about that we admire the demon, or *reasons why* our admiring the demon would be good, for intuitions about *reasons for* us to admire the demon.

So, there is a good explanation of why (A) seems so plausible but does not in fact hold, but there is no similar good explanation of why (B) or (C) seem so plausible but do not in fact hold. So, we should give up (A) and hold that there are no (non-elliptical) reasons to admire the demon for its own sake and desire the saucer of mud for its own sake when the demon will punish us if we do not. So, we should give up the idea that there are wrong kind of reasons for the pro-attitudes that figure in FA accounts. And since we should give up the idea that there are wrong kind of reasons for the pro-attitudes that figure in FA accounts, we should hold that FA accounts of value do not need to solve the WKR problem because there is no WKR problem.[33]

[32] See Dancy (2000, pp. 6–7).

[33] An anonymous referee claims that opponents of *FA* accounts may accept the argument of this paper but still claim that a slightly different wrong kind of reason problem arises for *FA* accounts. Suppose that an evil demon will severely punish us if we do not intrinsically desire a saucer of mud. In this case we have reason to intrinsically desire a saucer of mud. So, according to *FA* accounts, the saucer of mud is of intrinsic value. But the saucer of mud is not intrinsically valuable. So, *FA* accounts are implausible because they overgenerate intrinsic value.

However, it seems to me that *FA* accounts should not differentiate between *X* being of intrinsic or extrinsic value in terms of the pro-attitudes that there are reasons to have towards *X* but rather in terms of the properties that provide reasons for having these pro-attitudes in response to *X*. That is, proponents of *FA* accounts should hold that whether something is of intrinsic or extrinsic value depends on whether the reasons to have pro-attitudes towards it are provided by intrinsic or extrinsic properties of *X*. So, even if a demon's threat could give us reasons to intrinsically desire a saucer of mud, this would not show that *FA* accounts overgenerate intrinsic value, since *FA* accounts can and should give an account of the intrinsic/extrinsic value distinction other than in terms of intrinsic/non-intrinsic pro-attitudes.

Furthermore, it seems to me that there could be no reason to intrinsically desire a saucer of mud for the reason that a demon will punish you if you do not. This is because all the arguments that I gave in Sect. 2 regarding how we should understand claims about doing something 'for its own sake' have natural analogues regarding how we should understand claims about doing something 'intrinsically'. And so, it will be impossible to desire *X* intrinsically for reasons provided by non-intrinsic features of *X* for the same reasons that it is impossible to desire *X* for its own sake for the good consequences of desiring *X* for its own sake.

Acknowledgments I would like to thank Brad Hooker, Anthony Price, Philip Stratton-Lake, Bart Streumer, Jonathan Way, and several anonymous reviewers for extremely helpful comments on previous drafts of this paper. I would also like to thank audiences at University of Reading, the University of Sheffield's Understanding Value conference, and the Joint Session of the Aristotelian Society and Mind Association's Postgraduate Session.

References

Brentano, F. (1969). *The origin of our knowledge of right and wrong*. (Roderick Chisholm, Trans.). London: Routledge and Kagan Paul.

Buckley, C. (2007). Man is rescued by stranger on subway tracks. *The New York Times*. http://www.nytimes.com/2007/01/03/nyregion/03life.html.

Dancy, J. (2000). *Practical reality*. Oxford: Clarendon.

Daniellson, S., & Olson, J. (2007). Brentano and the buck-passers. *Mind, 116*, 511–522.

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford: Oxford University Press.

Ewing, A. C. (1947). *The definition of good*. London: Macmillan.

Heuer, U. (2011). Beyond wrong reasons: The buck-passing account of value. In M. Brady (Ed.), *New waves in metaethics*. Palgrave: Basingstoke.

Korsgaard, C. (1983). Two distinctions in goodness. *Philosophical Review, 92*(2), 169–195.

Lang, G. (2008). The right kind of solution to the wrong kind of reason problem. *Utilitas, 20*, 472–489.

Markovitz, J. (2011). Internal reasons and the motivating intuition. In M. Brady (Ed.), *New waves in metaethics*. London: Palgrave Macmillan.

McDowell, J. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity*. London: Routledge.

Olson, J. (2009). The wrong kind of solution to the wrong kind of reason problem. *Utilitas, 21*, 225–232.

Parfit, D. (2001). Rationality and reasons. In D. Egonsson, B. Petersson, J. Josefsson, & T. Rønnow-Rasmussen (Eds.), *Exploring practical philosophy: From action to values*. Aldershot: Ashgate.

Parfit, D. (2011). *On what matters* (Vol. 1). Oxford: Oxford University Press.

Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting attitudes and value. *Ethics, 114*(3), 391–423.

Rabinowicz, W., & Rønnow-Rasmussen, T. (2006). Buck-passing and the right kind of reasons. *Philosophical Quarterly, 56*, 114–120.

Raz, J. (2011). *From normativity to responsibility*. Oxford: Oxford University Press.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Schroeder, M. (2007). *Slaves of the passions*. Oxford: Oxford University Press.

Schroeder, M. (2010). Value and the right kind of reason. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 5). Oxford: Oxford University Press.

Schroeder, M. (2012). The ubiquity of state-given reasons. *Ethics, 122*, 457–488.

Shah, N. (2006). A new argument for evidentialism. *The Philosophical Quarterly, 56*, 481–498.

Skorupski, J. (2007). Buck-passing about goodness. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, & D. Egonsson (Eds.), Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.

Skorupski, J. (2010). *The domain of reasons*. Oxford: Oxford University Press.

Stratton-Lake, P. (2005). How to deal with evil demons: Comment on Rabinowicz and Rønnow-Rasmussen. *Ethics, 115*(4), 788–798.

Stratton-Lake, P., & Hooker, B. (2006). Scanlon versus Moore on goodness. In T. Horgan & M. Timmons (Eds.), *Metaethics after Moore*. Oxford: Oxford University Press.

Streumer, B. (2013). Can we believe the error theory? *Journal of Philosophy, 110*(4), 194–212.

Way, J. (2012). Transmission and the wrong kind of reason.*Ethics,122*, 489–515.

Way, J. (2013). Value and reasons to favour. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 8). Oxford: Oxford University Press.

Wiggins, D. (1987). A sensible subjectivism? In *Needs, values, truth: Essays in the philosophy of value*. Oxford: Blackwell.

Williams, B. (1989). Internal reasons and the obscurity of blame. In *Making sense of humanity and other philosophical papers*. Cambridge: Cambridge University Press.