

# Transmission and the Wrong Kind of Reason\*

Jonathan Way

University of Southampton

**ABSTRACT:** This paper defends fitting-attitudes accounts of value against the wrong kind of reason problem. I argue for the skeptical view that putative reasons of the wrong kind are reasons to want and bring about certain attitudes but not reasons for those attitudes. The argument turns on the transmission of reasons: the familiar fact that there is often reason for one action or attitude because there is reason for another. I argue that putative reasons of the wrong kind transmit in a different way to the right kind of reasons, and that this fact is best explained by the skeptical view.

## 1. Introduction

*Fitting-attitudes* accounts of value (FA accounts) hold, to a first approximation, that the valuable is what there is sufficient reason to value. Accounts of this form look especially plausible when applied to certain specific value properties. For example, each of the following analyses looks highly compelling:

For x to be admirable is for there to be sufficient reason to admire x.

For x to be enviable is for there to be sufficient reason to envy x.

For x to be amusing is for there to be sufficient reason to be amused by x.

These analyses are plausible because the admirable, the enviable, and the amusing are each conceptually connected to an attitude or reaction of a certain sort. But the connection is not a

---

\* For helpful comments on earlier versions of this paper, I would like to thank John Broome, Justin D'Arms, Guy Fletcher, Adrian Haddock, Errol Lord, Conor McHugh, Ryan Millsap, Jonas Olson, Sydney Penner, Andrew Reisner, Mark Schroeder, Jussi Suikkanen, Daniel Whiting, Fiona Woollard, a referee and several editors for *Ethics*, and audiences at the 3<sup>rd</sup> Rocky Mountain Ethics Congress at the University of Colorado, Boulder, the Oxford Moral Philosophy Seminar, and the Princeton Workshop on Normative Philosophy.

merely descriptive one – it is not that the admirable is what is *actually* admired, for instance. Rather, the connection is normative – the admirable is what *worth* admiring, or *fitting* to admire, or *correct* to admire. And if we understand this normative connection in terms of reasons – if we say that what is worthy, fitting, or correct is what there is sufficient reason for – then we reach the above analyses. By the same route, we reach parallel accounts of many other (dis)value properties, including the awesome, the scary, the blameworthy, the contemptible, the choiceworthy, the credible, and so on.<sup>1</sup>

However, FA accounts face a famous problem – the *wrong kind of reason* problem (WKR problem). The problem arises because it can seem that reasons to value something do not always bear on the object’s value. Suppose, for example, that an evil demon will kill you unless you admire him. Then, we might want to say, you ought to admire the demon, although he is not admirable. Or suppose, for a more realistic example, that envying your colleague’s recent promotion will be detrimental to your working relationship, and so to your own chances of promotion. Then, perhaps, you ought not envy your colleague. But this does not by itself show that your colleague’s promotion is not enviable.<sup>2</sup> FA accounts thus look subject to counter-example.

The WKR problem has a schematic solution. According to this solution, we need to distinguish between two kinds of reasons for attitudes. The *right* kind of reasons to value are considerations which are both reasons to value an object and features that make it valuable. The *wrong* kind of reasons are considerations which are reasons to value an object but not features that make it valuable. For example, the fact that Angie gives up so much of her time

---

<sup>1</sup> Cf. Justin D’Arms and Daniel Jacobson, ‘Sentiment and Value’, *Ethics* 110 (2000): 722-48, 747 and Mark Schroeder, ‘Value and the Right Kind of Reason’, *Oxford Studies in Metaethics, Volume 5*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2010), 25-55, 26-7.

<sup>2</sup> The first of these examples is from Wlodek Rabinowicz and Toni Rønnow-Rasmussen, ‘The Strike of the Demon: On Fitting Pro-Attitudes and Value’, *Ethics* 114 (2004): 391-423, the second is from Justin D’Arms and Daniel Jacobson, ‘The Moralistic Fallacy: On the “Appropriateness” of Emotions’, *Philosophy and Phenomenological Research* 61 (2000): 65-90. For further examples and discussion, see the works cited in note 3.

to charity work is both a reason to admire her and something that makes her admirable. By contrast, the fact that the demon will kill you unless you admire him is a reason to admire the demon but not a feature that makes him admirable. The task for FA analyses is to draw this distinction in an informative way, and without presupposing notions of value. If this task can be completed, FA analyses can be revised: we can say that the valuable is what there is sufficient reason of the right kind to value.<sup>3</sup>

As many have noted, the distinction FA accounts require here seems to be of a kind with the familiar distinctions between *evidential* and *pragmatic* reasons for belief, and between reasons to intend which depend on reasons to act and reasons to intend which depend on further benefits of intending. The former distinction is famously illustrated by Pascal's wager. According to Pascal, the expected utility of believing in God is a reason to believe in God, although it is not evidence that God exists. The latter distinction is well-illustrated by Kavka's toxin puzzle.<sup>4</sup> Kavka imagined an eccentric billionaire offering you a \$1m prize for intending today to drink an unpleasant toxin at noon tomorrow. Here there appears to be a reason to intend to drink the toxin, but – since you win the money whether or not you go on to drink the toxin – not a reason to drink it.

It is immediately plausible that the same sort of thing is going on in all of these cases – that what we see here are three instances of a general distinction. This immediate reaction can be reinforced by observing that there are common marks shared by the contrasting reasons in these cases. Most strikingly, there is a difference in the ease and way in which you

---

<sup>3</sup> For attempts to carry out this task, see D'Arms and Jacobson 'Sentiment and Value' and 'The Moralistic Fallacy'; Sven Danielsson and Jonas Olson, 'Brentano and the Buck-Passers', *Mind* 116 (2007): 511-22; Pamela Hieronymi, 'The Wrong Kind of Reason', *Journal of Philosophy* 102 (2005): 437-57; Jonas Olson, 'Buck-Passing and the Wrong Kind of Reasons', *The Philosophical Quarterly* 54 (2004): 295-300; Derek Parfit, 'Reasons and Rationality', in *Exploring Practical Philosophy*, ed. Dan Egonsson, Bjorn Petersson, Jonas Josefsson, and Toni Rønnow-Rasmussen, *Exploring Practical Philosophy: From Action to Values* (Aldershot: Ashgate, 2001), 17-39; Christian Piller, 'Content-Related and Attitude-Related Reasons for Preferences', *Philosophy* 59 (2006): 151-81; Rabinowicz and Rønnow-Rasmussen, 'The Strike of the Demon'; Schroeder, 'Value and the Right Kind of Reason'; Phillip Stratton-Lake, 'How to Deal with Evil Demons: Comment on Rabinowicz and Rønnow-Rasmussen', *Ethics* 115 (2005): 788-98.

<sup>4</sup> Gregory Kavka, 'The Toxin Puzzle', *Analysis* 43 (1983): 33-6.

can respond to the different reasons. For example, it is hard to see how you could believe in God on grounds of expected utility, intend to drink the toxin in order to win \$1m, or admire the demon on account of his threat. By contrast, believing on account of the evidence, intending on the basis of reasons to act, and admiring the admirable, are familiar, if poorly understood, phenomena. Moreover, there is a difference in the kind of overall evaluation on which the different reasons bear. The incentives in our examples perhaps make it the case that you *ought* to believe in God, intend to drink the toxin, and admire the demon. But they do not make these attitudes *correct* or *fitting*. It seems to be correct to believe only what is true, admire only the admirable, and intend to do only what is worth doing. These shared marks strongly suggest a general distinction between two kinds of reasons for attitudes. If this is right, an important constraint on solutions to the WKR problem is that evidential reasons to believe and ‘act-based’ reasons to intend are counted as being of the right kind. In what follows, I thus use the terminology of the ‘right’ and ‘wrong’ kind of reasons broadly, to mark this general distinction, of which the distinction between the two kinds of reason to value is an instance.

This paper defends a very simple solution to the WKR problem, understood in this broad way. According to this solution, the right kind of reasons are simply all the reasons there are. Strictly speaking, the wrong kind of reasons for attitudes are not reasons for those attitudes, any more than fool’s gold is gold. The demon’s threat is not a reason to admire the demon, expected utility is not a reason to believe in God, and the \$1m prize is not a reason to intend to drink the toxin. Rather, the incentives in these cases are reasons to *want* these attitudes, and to bring it about that you have these attitudes, if you can. More generally, incentives for attitudes are not reasons for those attitudes, but are instead reasons to want these attitudes, and to bring them about.

This view treats incentives for attitudes like admiration, belief, and intention in the way it is natural to treat incentives for states for which it is clear that there cannot be reasons. Just as a demon might threaten to kill you unless you admire him, a demon might threaten to kill you unless you have a headache, or an experience as of a red cube, or unless your hair is turquoise. Since headaches, perceptual experiences, and the colour of your hair are not the sorts of things there can be reasons for, such threats would not provide reasons for you to have a headache, an experience of a red cube, or for your hair to be turquoise. But these threats would give you reasons to want these things, and to bring them about. On the view defended here, the incentives in *wrong kind of reason scenarios* should be treated in the same way. And when treated in this way, the WKR problem disappears. I call this view *WKR skepticism*.

WKR skepticism is not new. It has been suggested by Alan Gibbard, Derek Parfit, and John Skorupski, among others.<sup>5</sup> However, many proponents of the WKR problem – *WKR defenders* – find the view unsatisfactory. Thus Rabinowicz and Rønnow-Rasmussen complain that ‘Gibbard and Parfit have no argument for their claim, apart from an appeal to intuitions we do not share’.<sup>6</sup> And Sven Danielsson and Jonas Olson agree:

---

<sup>5</sup> See Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, Mass: Harvard University Press, 1990); Thomas Kelly, ‘The Rationality of Belief and Some Other Propositional Attitudes’, *Philosophical Studies* 90 (2002): 163-96; Niko Kolodny, ‘Why be Rational?’, *Mind* 114 (2005): 509-63; David Owens, *Reason without Freedom* (Oxford: Routledge, 2000); Parfit, ‘Reasons and Rationality’ and *On What Matters* (Oxford: Oxford University Press, 2011), chap. 2 and appendix A; Ingmar Persson, ‘Primary and Secondary Reasons’, in *Homage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, ed. Toni Rønnow-Rasmussen, Bjorn Petersson, Jonas Josefsson and Dan Egonsson, online resource, <http://www.fil.lu.se/hommageawlodek/>; Thomas Pink, *The Psychology of Freedom* (Cambridge: Cambridge University Press, 1996); Nishi Shah, ‘A New Argument for Evidentialism’, *The Philosophical Quarterly* 56 (2006): 481-98, and ‘How Action Governs Intention’, *Philosophers’ Imprint* 8 (2008); John Skorupski, *The Domain of Reasons* (Oxford: Oxford University Press, 2010). Not all of these authors are committed to WKR skepticism about reasons for all of the attitudes we are concerned with here.

<sup>6</sup> ‘The Strike of the Demon’, 412.

[i]t is difficult to find an independent rationale for distinguishing in the relevant way between reasons for an attitude and reasons to bring it about that we have that attitude...in the demon scenario, the intuition that we do have reasons to favour the demon seems at least as strong as any contrary intuition.<sup>7</sup>

It is perhaps unfair to say that WKR skeptics have offered no argument for their claim. Several authors argue for WKR skepticism by appealing to one of the marks of the wrong kind of reason noted above – that we seem not to be able to respond to the wrong kind of reasons by forming the relevant attitude. These authors argue that we cannot respond to such reasons in this way, and that if we cannot respond to some consideration by forming an attitude, then that consideration is not a reason for that attitude. However, while this argument has significant appeal, it is difficult to get the details straight – to specify a notion of ‘responding to a reason’ that will vindicate its two premises.<sup>8</sup> So it is worth considering whether WKR skepticism might be defended on other grounds.

In this paper, I offer a new argument for WKR skepticism – a new way of meeting the challenge posed by Rabinowicz and Rønnow-Rasmussen, and by Danielsson and Olson. The argument I offer turns on the familiar fact that reasons *transmit*: that there is often reason for one action or attitude because there is reason for another. For example, there is often reason to do one thing because it is a means to doing something else there is reason to do. And there is often reason to believe something because it follows from something else there is reason to believe. I shall begin by detailing some of the ways in which reasons transmit, and arguing that putative reasons of the wrong kind transmit in a very different way to reasons of the right kind. This difference, I suggest, calls for explanation. We should expect an account of

---

<sup>7</sup> ‘Brentano and the Buck-Passers’, 513-4.

<sup>8</sup> For versions of this argument, see Kelly, ‘The Rationality of Belief’, Kolodny, ‘Why be Rational?’, 547-51, Persson, ‘Primary and Secondary Reasons’, and Shah, ‘A New Argument for Evidentialism’. For discussion and some worries, see Hieronymi, ‘The Wrong Kind of Reason’, Jennie Louise, ‘Correct Responses and the Priority of the Normative’, *Ethical Theory and Moral Practice* 12 (2009): 345-64, and Andrew Reisner, ‘The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem’, *Philosophical Studies* 145 (2009): 257-72.

reasons of the wrong kind to explain the way in which such reasons behave. The central argument of this paper is that WKR skeptics succeed in meeting this challenge, and WKR defenders fail. In section three, I argue that WKR skeptics can explain why putative reasons of the wrong kind transmit in the way that they do, by appealing to the ways in which reasons to want and bring about attitudes transmit. In section four, I consider two strategies by which WKR defenders might try to explain the way in which reasons of the wrong kind transmit. I argue that both these strategies fail, unless they are supplemented with a further thesis. But this further thesis, I argue, is false. WKR defenders are thus left without a way to explain the transmission of the wrong kind of reasons. I conclude that the difference in the ways in which reasons of the right and wrong kind transmit provides important support for WKR skepticism.

## **2. Transmission Patterns**

It will be helpful to begin by looking in some detail at the way in which reasons for action transmit from ends to means.

Suppose that there are reasons for you to visit some friends in another part of the world – for example, in California. Because of this, there are reasons for you to take means to going to California – to book a ticket, travel to the airport, board the plane, and so forth. There are also reasons for you to ensure that certain enabling conditions of your going to California are fulfilled – you might need to organize time off work, or renew your passport. In other cases, there are reasons to take so called ‘constitutive means’ – reasons to swim as a way of exercising, or to go to the concert as a way of having a pleasant evening. Following Joseph Raz, I will say that any action that helps you to achieve an end (in one of these ways,

or perhaps others) *facilitates* that end.<sup>9</sup> The examples suggest that we can say that reasons for action transmit in the following way:

(Action Pattern)      If there is a reason to A, then the fact that B-ing facilitates A-ing is a reason to B.

Several points of clarification and qualification are in order.<sup>10</sup>

First, the Action Pattern says not merely that there is a reason to B, when B-ing facilitates A-ing; it says that the fact that B-ing facilitates A-ing *is* a reason to B. This gives us a weak sense in which there is a reason to B *because* there is a reason to A. Specifying the reason in this way will also make it easier to assess some of the claims about reasons made below.

Second, some may think that the fact that B-ing facilitates A-ing cannot itself be a reason to B, even when there is reason to A. Rather, the reasons to B, fully spelled out, will be the facts that constitute reasons to A, in conjunction with the fact that B-ing facilitates A-ing. Readers of this view will prefer to formulate the Action Pattern as follows: if R is a reason to A, then the fact that R and that B-ing facilitates A-ing is a reason to B. I believe that nothing in what follows turns on the difference between this formulation and that above. However, for brevity I will stick with the original formulation.

Third, if you are certain not to achieve the end, whether or not you take the means, we might doubt that the end provides any reason to take the means. If you have resolved not to go to California, even if you have a ticket, we may want to deny that the fact that buying a ticket could help you get to California is a reason to buy a ticket. We can account for this by

---

<sup>9</sup> Joseph Raz, 'The Myth of Instrumental Rationality', *Journal of Ethics and Social Philosophy* 1 (2005): 1-28.

<sup>10</sup> Note also that a more precise version of the Action Pattern would include variables for the agent, and the times at which the reason applies, and at which A-ing is to be done. For brevity, I ignore these details throughout.

stipulation: we will say that B-ing facilitates A-ing only if there is some chance that you will A, conditional on B-ing.<sup>11</sup>

Fourth, if you are certain *to* achieve the end, whether or not you take a certain means, we might also doubt that the end provides any reason for this means. Thus if you already have a ticket for California, the fact that buying a further ticket could also help you get to California seems to be no reason at all to buy another ticket. We can account for this by further specification: it is only if B-ing *non-superfluously* facilitates A-ing, that reasons to A transmit to reasons to B.<sup>12</sup>

Fifth, the Action Pattern makes no claim about the *weight* of reasons. A natural suggestion, following Mark Schroeder and Niko Kolodny, is to say that *how much* reason is transmitted from A to B is proportional to *how well* B-ing facilitates A-ing.<sup>13</sup> However, nothing in what follows turns on the weight of reasons, so I leave the issue open.<sup>14</sup>

---

<sup>11</sup> See Niko Kolodny, 'Instrumental Reasons' (unpublished manuscript, University of California, Berkeley, 2011). I leave the relevant notion of probability open. The reader will notice that the relation of facilitation across which reasons for action transmit is very weak. For some reasons to think the relevant relation needs to be very weak, see Mark Schroeder, *Slaves of the Passions* (Oxford: Oxford University Press, 2007), 112-3 and Kolodny, 'Instrumental Reasons'.

<sup>12</sup> Kolodny, 'Instrumental Reasons'.

<sup>13</sup> Mark Schroeder, 'Means-End Coherence, Stringency, and Subjective Reasons', *Philosophical Studies* 43 (2009): 223-48, 246.

<sup>14</sup> The Action Pattern raises two further issues which I leave open. First, the pattern may seem to have implausible consequences in cases of objectionable means. Suppose, for example, that you have a reason to get the job, and that secretly assassinating your rival facilitates your doing so. The Action Pattern implies that this latter fact is a reason to assassinate your rival. Some will find this objectionable (cf. John Broome, 'Have We Reason To Do As Rationality Requires? A Comment on Raz', *Journal of Ethics and Social Philosophy* 1 (2005): 1-9, 7). A reasonable response, it seems to me, is to accept the implication, but deny that it is objectionable. It sounds implausible to say that there is a reason to assassinate your rival only because assertions of this form carry a pragmatic implication that this reason is relatively weighty (cf. Donald Hubin, 'What's Special About Humeanism,' *Noûs*, 33 (1999): 30-45, 34; Schroeder, *Slaves of the Passions*, 92-97; and Kolodny, 'Instrumental Reasons'). However, those not convinced by this response may prefer to qualify the Action Pattern so as to avoid counter-examples of this sort.

Second, the Action Pattern faces what Matthew Bedke calls the 'problem of subversion' ('The Iffiest Oughts: A Guise of Reasons Account of End-Given Conditionals', *Ethics* 119 (2009): 672-98, n.12). Suppose, to borrow an example from Ryan Millsap, that you have reason to play tennis, and that buying tennis balls facilitates playing tennis. The Action Pattern implies that this fact is a reason to buy tennis balls. But if selling your tennis racket facilitates buying tennis balls, then the Action Pattern also implies that this fact is a reason to sell your tennis racket. And assuming you need your tennis racket to play tennis, this seems pretty implausible. This problem deserves more attention than I can give it here – see Bedke, 'The Iffiest Oughts', n.12, Kolodny, 'Instrumental Reasons', and Millsap, 'Practical Reasons and Means-End Transmission' (PhD Dissertation, University of Maryland, in progress) for discussion. One point to note is that the problem does not arise for a version of the Action Pattern restricted to *necessary* facilitative steps. Much of the discussion to follow could be framed in these terms.

With the Action Pattern in hand, we can make our first claim about transmission for the right kind of reasons. Since the right kind of reasons to intend depend on reasons to act, we should expect the right kind of reasons to intend to exhibit a pattern corresponding to the Action Pattern. And that is what we find: when there are reasons of the right kind to intend to go to California, the fact that booking a ticket facilitates going to California is a reason of the right kind to intend to book a ticket, the fact that renewing your passport facilitates going to California is a reason of the right kind to intend to renew your passport, and so on. Thus, and subject to the same qualifications as the Action Pattern<sup>15</sup>, we can say that:

(Intention Pattern)     If there is a reason of the right kind to intend to A, then the fact that B-ing facilitates A-ing is a reason of the right kind to intend to B.

Consider now the right kind of reasons to desire. Such reasons transmit in a similar way to reasons to intend. If there are reasons of the right kind to want the conference you are attending next week to go well, there are reasons of the right kind to want things that will help to bring this about, or which are conditions of its doing so – the fact that good papers, good weather, and good conversation will help the conference to go well is a reason to want good papers, good weather, and good conversation. The right kind of reasons to desire transmit across facilitative connections between the objects of desire:

(Desire Pattern)        If there is a reason of the right kind for you to desire x to F, then the fact that y's G-ing facilitates x's F-ing is a reason of the right kind for you to desire y to G.<sup>16</sup>

---

<sup>15</sup> The Intention Pattern is also subject to a further qualification: if *intending* to B does not facilitate *B-ing*, reasons to B do not provide reasons to intend to B.

<sup>16</sup> The greater complexity in this formulation is required to cover the case in which, for example, there is a reason to want the sun to shine because the sun's shining facilitates your garden's growing.

We can now turn to putative reasons of the wrong kind. The first thing to note is that the wrong kind of reasons do not transmit in the way that the right kind of reasons do – the patterns we get by replacing ‘right’ with ‘wrong’ in the Intention and Desire Patterns are false. To see this, suppose that in the toxin puzzle, going to the billionaire’s mansion for noon the next day facilitates drinking the toxin. The Intention Pattern, with ‘right’ replaced by ‘wrong’, tells us that this fact is a reason of the wrong kind to intend to go to the billionaire’s mansion. But although this fact might be a reason of the right kind to intend to go to the billionaire’s mansion – if, for instance, there is any reason at all to drink the toxin – it is not a reason of the wrong kind for this intention. Similarly, if an evil demon threatens to kill you unless you desire a saucer of mud, then the fact that acquiring a saucer and heading out into the garden facilitates getting a saucer of mud might be a reason of the right kind to want to do this – if there is anything desirable about having a saucer of mud – but it is not a reason of the wrong kind for this desire.

However, putative reasons of the wrong kind do transmit in other ways. Here are two examples. First, if you believed, in the toxin puzzle, that you ought to drink the toxin, then you could form the intention to drink the toxin on that basis. (You could simply reason, ‘I ought to drink the toxin, so I’ll drink the toxin’, thereby forming the intention to do so). In this way, believing that you ought to drink the toxin would help you to form the intention to drink the toxin, and so help you to win the \$1m. That looks like a reason to believe that you ought to drink the toxin. However, it is not evidence that you ought to drink the toxin. Rather, it is a way in which it would be good to believe that you ought to drink the toxin. It is a reason of the wrong kind for this belief.

Second, suppose that you believe that going to the billionaire’s mansion is necessary for drinking the toxin. If you are rational, you will find it difficult to intend to drink the toxin

unless you also intend to go to the billionaire's mansion – for if you are rational, you will intend what you take to be the necessary means to your intended ends. In this way, intending to go to the billionaire's mansion would help you to intend to drink the toxin, and so win the \$1m. That looks like a way in which it would be good to intend to go to the billionaire's mansion, and so a reason of the wrong kind for this intention.

In the first of these examples, the belief that you ought to drink the toxin is a means to intending to drink the toxin. In the second, the intention to go to the billionaire's mansion is an enabling condition. So what the examples suggest is that putative reasons of the wrong kind transmit across facilitative connections between attitudes – that when one attitude facilitates an attitude for which there is the wrong kind of reason, the fact that it does so is a reason of the wrong kind for that attitude. Again, this claim needs to be qualified in various ways. For example, if you have just taken an “intention pill”, which will ensure that at midnight tonight you will intend to drink the toxin, then the belief that you ought to drink the toxin is superfluous. So in this version of the case, there is no benefit to, and so no reason of the wrong kind of reason for, this belief. For another example, if you are unable to sustain an intention to drink the toxin even if you intend to go to the billionaire's mansion, then forming this latter intention is pointless in a different way – in the sense of ‘facilitates’ stipulated earlier, it does not facilitate the intention to drink the toxin. So in this case too, there is no benefit to, and so no reason of the wrong kind for, this intention.

Keeping these qualifications in mind however – and noting in passing that they parallel the qualifications made to the Action Pattern – the examples nonetheless suggest that putative reasons of the wrong kind transmit in the following way:

(Wrong Reason Pattern)                      If there is a reason of the wrong kind for attitude A,  
then the fact that attitude B facilitates attitude A is a

reason of the wrong kind for attitude B.

I now want to suggest that this pattern is distinctive of the wrong kind of reasons for attitudes. No parallel pattern holds for the right kind of reasons, or for reasons for attitudes in general.

Consider first the pattern we get by replacing both occurrences of ‘wrong’ with ‘right’:

(Right Reason Pattern)                      If there is a reason of the right kind for attitude A, then the fact that attitude B facilitates attitude A is a reason of the right kind for attitude B.

This pattern clearly does not hold. For example, consider someone who has evidence that  $p$ , but is disposed to believe that  $p$  only if he also believes that  $q$ . And suppose that this disposition holds not because of any evidential relation between  $p$  and  $q$ , but because of a strange psychological quirk, or because believing that  $q$  helps him to cope with the painful truth that  $p$ . For such a person, believing that  $q$  facilitates believing that  $p$ . But that fact is not an evidential reason to believe that  $q$ . The right kind of reasons to desire also illustrate the point. Consider someone who has a reason of the right kind to desire to A, but is psychologically disposed to desire to A only if she also desires to B. For such a person, desiring to B facilitates desiring to A. Again though, that fact is not a reason of the right kind to desire to B.<sup>17</sup>

Consider now the:

---

<sup>17</sup> Cf. Kieran Setiya, ‘Cognitivism about Instrumental Reason’, *Ethics* 117 (2007): 649-73, 672 and Jonathan Way, ‘Defending the Wide-Scope Approach to Instrumental Reason’, *Philosophical Studies* 147 (2010): 213-33, 226-7.

(General Pattern)                      If there is a reason for attitude A, then the fact that attitude B facilitates attitude A is a reason of the wrong kind for attitude B.

The General Pattern combines two claims. The first is the Wrong Reason Pattern. The second is the:

(Mixed Pattern)                      If there is a reason of the right kind for attitude A, then the fact that attitude B facilitates attitude A is a reason of the wrong kind for attitude B.

This pattern may seem plausible. For it may seem that it is good to have attitudes there is the right kind of reason for. Of course, such attitudes need not be instrumentally good. But such attitudes may be thought to be good for their own sake, insofar as they constitute a proper response to one's reasons. If so, then attitudes which facilitate attitudes there is the right kind of reason for are instrumentally good, and so attitudes there is the wrong kind of reason for.

However, this argument for the Mixed Pattern fails. It may well be good for its own sake to have attitudes which constitute a proper response to one's reasons. But we do not properly respond to our reasons by having attitudes there is merely *some* reason for. We properly respond to our reasons only when we have attitudes there is *sufficient* reason for. So the appeal to value of this sort does not support the claim that it is good for its own sake to have attitudes there is reason of the right kind for.<sup>18</sup>

---

<sup>18</sup> Even the claim that it is good to have attitudes there is sufficient reason for is controversial. For example, it is denied by Joseph Raz ('Reasons: Practical and Adaptive', in *Reasons for Action*, ed. David Sobel and Steven Wall (Cambridge: Cambridge University Press, 2009), 43, 47-8) and Ulrike Heuer ('Beyond Wrong Reasons: The Buck-Passing Account of Value', in *New Waves in Metaethics*, ed. Michael Brady (Basingstoke: Palgrave Macmillan), 166-84, 177-8), and it is likely that it would be denied by many of the philosophers who hold that there is no distinctive value in believing the truth. For some discussion of the final value of conative attitudes,

Furthermore, examples suggest that this claim is false. For example, if Mustard had access to the murder weapon, then that is some reason to think he committed the crime. But if several others also had access to the murder weapon, and we have no further relevant evidence, then there is not sufficient reason to believe that Mustard committed the crime. If you nonetheless do believe that Mustard committed the crime, you are not responding properly to your reasons, and your belief does not seem to be good for its own sake. Nor does it seem good for its own sake to desire what there is merely some reason to desire, when this reason is not sufficient.<sup>19</sup>

If it is not necessarily good to have attitudes there is the right kind of reason for, then there is not necessarily anything instrumentally good about attitudes which facilitate attitudes there is the right kind of reason for. And so there is not necessarily reason of the wrong kind for such attitudes. We should therefore reject the Mixed Pattern, and so the General Pattern. The Wrong Reason Pattern is a distinctive feature of the wrong kind of reasons.

### **3. WKR Skepticism and the Wrong Reason Pattern**

The ways in which different kinds of reasons transmit should not, I suggest, be taken as primitive. When reasons of a certain kind have some distinctive feature, we should expect an account of reasons of that kind to explain that feature. Thus we should expect an account of the wrong kind of reasons to explain the ways in which such reasons transmit.

This task takes slightly different forms for WKR skeptics and WKR defenders. WKR defenders must explain why the wrong kind of reasons do in fact transmit in accordance with the Wrong Reason Pattern. WKR skeptics, since they deny that there are any reasons of the wrong kind, must instead explain why it might appear that the wrong kind of reasons transmit

---

see Thomas Hurka, *Virtue, Vice, and Value* (Oxford: Oxford University Press, 2003) and my 'Value and Reasons to Favour' (unpublished manuscript, University of Southampton, 2011).

<sup>19</sup> I defend these claims further in 'Value and Reasons to Favour'.

in this way. They must explain why, if we take reasons to want and bring about attitudes for reasons for those attitudes, we will also take the Wrong Reason Pattern to hold. In this section, I shall argue that WKR skeptics can meet this challenge, by appealing to the ways in which reasons for desire and action transmit.

The Wrong Reason Pattern is strikingly similar to the Action Pattern. Where the latter says that there is reason to do what facilitates actions there is reason to do, the former says that there is reason for attitudes which facilitate attitudes there is reason (of the wrong kind) to have. In both cases, reasons transmit across facilitative connections between the things for which there are reasons.<sup>20</sup>

The connection between the Action and Wrong Reason Pattern becomes clearer still when we consider the following near-instance of the Action Pattern:

(Bringing About Pattern):

If there is a reason to bring about attitude A, then the fact that attitude B facilitates attitude A is a reason to bring about attitude B.

The Bringing About Pattern is not quite an instance of the Action Pattern. Strictly speaking, the Action Pattern only tells us that when there is reason to bring about an attitude A, the fact that *bringing about* attitude B facilitates *bringing about* attitude A is a reason to bring about attitude A. But at least typically, bringing about an attitude B facilitates bringing about attitude A only because *having* attitude B facilitates *having* attitude A. So it is plausible that

---

<sup>20</sup> Of course, the right kind of reasons for intention and desire also transmit across facilitative connections. But the facilitative connections in question do not hold between the things for which there are reasons. Rather, they hold between the *objects* of the things for which there are reasons – there is reason to want *good papers* because there is reason to want *the conference to go well*.

when a reason to bring about attitude B is that doing so facilitates bringing about attitude A, the fact that attitude B facilitates attitude A will also be a reason to bring about attitude B.

The Bringing About Pattern is exactly parallel to the Wrong Reason Pattern. We can move from the former to the latter by replacing ‘reason to bring about’ with ‘reason of the wrong kind for’. So if we take reasons to bring about an attitude to be reasons for that attitude, we will take the Wrong Reason Pattern to hold, given that the Bringing About Pattern holds.

The Wrong Reason Pattern is related to the Desire Pattern in a very similar way. Consider the following instance of the Desire Pattern:

(Instance of the Desire Pattern)      If there is a reason of the right kind to want attitude A, then the fact that attitude B facilitates attitude A is a reason of the right kind to want attitude B.

Again, replacing ‘reason of the right kind to want’ with ‘reason of the wrong kind for’ takes us from the Instance of the Desire Pattern to the Wrong Reason Pattern. So if we take reasons to desire an attitude to be reasons for that attitude, we will take the Wrong Reason Pattern to hold, given that the Instance of the Desire Pattern holds.

WKR skeptics can thus offer a straightforward explanation of why the Wrong Reason Pattern appears to hold. This explanation offers some support for WKR skepticism. But how much support it offers depends on whether WKR defenders can adequately explain the Wrong Reason Pattern. It is to this question I now turn.

#### **4. WKR Defenders and the Wrong Reason Pattern**

In this section, I shall consider two strategies by which WKR defenders can try to explain the Wrong Reason Pattern. The first looks to WKR defenders' own accounts of the distinction between the right and the wrong kind of reasons. The second considers the possibility of analysing the wrong kind of reasons in terms of value. I shall argue that in and of themselves, these approaches fail. The first simply fails to offer any explanation of the Wrong Reason Pattern. The second seems incompatible with FA accounts of value. However, I shall also argue that these strategies are more successful when supplemented with a thesis I call Reductionism. Reductionism holds that the wrong kinds of reasons for an attitude are to be analysed in terms of the right kind of reasons to want that attitude. Given this thesis, the two strategies turn out to be equivalent – and successful. However, I shall argue that Reductionism is not a tenable analysis of reasons of the wrong kind. If this is right, WKR defenders are left without a satisfactory explanation of the Wrong Reason Pattern.

#### **4.1. First Strategy: Appealing to Existing Theories of the Right/Wrong Distinction**

It is natural to think that to explain the Wrong Reason Pattern, WKR defenders should look to their own accounts of the distinction between the right and wrong kind of reasons. But it turns out that this is not an especially promising strategy. To explain the Wrong Reason Pattern, WKR defenders need to give an account of what reasons of the wrong kind are, and show that it follows from this account that reasons of the wrong kind transmit across facilitative connections. But most accounts of the distinction between reasons of the right and wrong kind are in no position to do this. Most accounts are primarily theories of the *right* kind of reasons, and tell us little about the wrong kind of reasons. This is not, in itself, a problem for such accounts. If the aim is simply to defend a FA account of value, it is perfectly legitimate to proceed by telling us what the right kind of reasons are, and then

framing your FA account in those terms. However, it does mean that such accounts are in no position to show that the wrong kind of reasons transmit across facilitative connections, and so offer no explanation of the Wrong Reason Pattern. I will illustrate this point by looking at some examples.

The best-known account of the distinction between the right and the wrong kind of reasons is what I will call the ‘object/state’ theory. On this view, the right kind of reasons for an attitude depend on properties of the object of the attitude; the wrong kind of reasons depend on properties of the attitude itself.<sup>21</sup> For example, Angie’s generosity is the right kind of reason to admire her, because it depends on a property of Angie – her generosity. But the fact that admiring the demon saves your life is the wrong kind of reason to admire him, since it depends on a property of the admiring itself – that it will save your life.

This view does tell us something about reasons of the wrong kind. But what it tells us is not sufficient to explain the Wrong Reason Pattern. To explain the Wrong Reason Pattern, proponents of the object/state theory would have to do two things. First, they would have to specify the properties which provide state-given reasons. Second, they would have to show that these properties transmit across facilitative connections. They would have to show that whenever an attitude B facilitates an attitude A, attitude B has state-given reason-providing properties if attitude A does. But proponents of the object/state theory have not done the first of these things. They have told us that state-given reasons depend on properties of the attitude, but they have not told us *which* such properties are relevant. And so proponents of this theory do not tell us enough to show that state-given reason-providing properties transmit across facilitative connections. They do not tell us enough to explain the Wrong Reason Pattern.

---

<sup>21</sup> See Olson, ‘Buck-Passing and the Wrong Kind of Reasons’, Parfit, ‘Reasons and Rationality’ and *On What Matters*, Piller ‘Content-Related and Attitude-Related Reasons’, and Stratton-Lake, ‘How to Deal with Evil Demons’.

A similar difficulty faces the ‘Brentano-style’ account offered by Danielsson and Olson.<sup>22</sup> Danielsson and Olson distinguish the right kind of reasons by appeal to the second mark of the right kind of reason noted in section one: that the right kind of reasons bear on correctness. Thus Danielsson and Olson tell us that the right kind of reasons are ‘content-reasons’: reasons that count in favour of an attitude’s correctness.<sup>23</sup> Angie’s generosity is the right kind of reason to admire her, because it counts in favour of its being correct to admire her; the demon’s threat is not a reason to admire him, because it does not count in favour of – indeed, it counts against – its being correct to admire him. Danielsson and Olson contrast content-reasons with ‘holding-reasons’. To explain the Wrong Reason Pattern, Danielsson and Olson would thus need to show that holding-reasons transmit across facilitative connections. They would need to show that whenever attitude B facilitates attitude A, the fact that it does so is a holding reason for attitude B if there is a holding reason for attitude A.<sup>24</sup> But all that Danielsson and Olson tell us about holding-reasons is that they are “reasons for having the attitude”.<sup>25</sup> So they do not tell us enough to explain the Wrong Reason Pattern.

Other views tell us even less about reasons of the wrong kind. For example, D’Arms and Jacobson suggest that the right kind of reasons for an attitude are those that bear on the truth of the content of the attitude (they focus on the right kind of reasons for emotion, although it is clear that the suggestion can also be applied to reasons for belief, and perhaps intention too).<sup>26</sup> But although they supply us with a wide-range of examples of reasons of the wrong kind, they make no general suggestion about what such reasons are. Similarly, Mark

---

<sup>22</sup> ‘Brentano and the Buck-Passers’.

<sup>23</sup> What Danielsson and Olson actually say is that content-reasons are ‘reasons for the correctness of an attitude’ (ibid., 515). It is a little difficult to interpret this. In particular, it is not clear whether they intend the relationship between content-reasons and correctness to be explanatory or evidential. I intend my use of ‘counts in favour of’ to be neutral between these interpretations.

<sup>24</sup> Things are actually a bit more complicated than this, since Danielsson and Olson’s view is that the wrong kind of reasons are a subset of holding-reasons: they are holding-reasons that do not derive from content-reasons (ibid., 517).

<sup>25</sup> Ibid., 515. Olson and Danielsson go on to offer an analysis of holding-reasons in terms of content-reasons. I address this suggestion below.

<sup>26</sup> ‘The Moralistic Fallacy’.

Schroeder defends a very interesting account on which the right kind of reasons are those shared by every agent engaged in a certain kind of activity (Schroeder understands this notion broadly, so as to include things like believing, admiring, and intending).<sup>27</sup> But Schroeder's view is also entirely silent about reasons of the wrong kind. On both of these views, reasons of the wrong kind are just left to be any reasons that are not of the right kind. Now as I have said, this is not by itself a problem for such accounts. But it does mean that these accounts cannot tell us whether reasons of the wrong kind transmit across facilitative connections, and so can offer no explanation of the Wrong Reason Pattern.

We have now seen why standard accounts of the distinction between the right and the wrong kind of reasons do not, in and of themselves, explain the Wrong Reason Pattern. I shall now show that an explanation becomes available if proponents of these accounts adopt a further thesis. This further thesis is a generalisation of a claim made explicitly by Danielsson and Olson. Having distinguished between content-reasons and holding-reasons, Danielsson and Olson go on to suggest that:

[T]he notion of a holding-reason should be analysed in terms of the notion of a content-reason. To say that there is a holding reason to have some attitude is to say that there is a content-reason to favour the occurrence of this attitude.<sup>28</sup>

On Danielsson and Olson's account, holding-reasons are reduced to content-reasons: to have a holding-reason for an attitude is to have a content-reason to favour that attitude. The other accounts we have considered may make a parallel move. Given an account of the right kind of reasons, the wrong kind of reasons may be analysed in terms of the right kind of reasons to want:

---

<sup>27</sup> 'Value and the Right Kind of Reason'.

<sup>28</sup> 'Brentano and the Buck-Passers', 518-9.

(Reductionism) For  $p$  to be a reason of the wrong kind for attitude A is for  $p$  to be a reason of the right kind to want attitude A.

Reductionists can explain the Wrong Reason Pattern. As we saw in section three, the wrong kind of reasons transmit in the same way as the right kind of reasons for higher-order desires. Reductionists can hold that this is so because what it is for some consideration to be a reason of the wrong kind just is for it to be a reason of the right kind for a higher-order desire.

#### 4.2. Second Strategy: The Value-Based Theory

I now turn to a second strategy by which WKR defenders might try to explain the Wrong Reason Pattern. I shall argue that this strategy also fails, unless it is supplemented by Reductionism.

The wrong kind of reasons for an attitude seem to have something important to do with the value of that attitude. When the demon threatens to kill you unless you admire him, it would be good to admire the demon, since this would help you to stay alive. When the billionaire offers you a \$1m to intend to drink the toxin, it would be good to intend to drink the toxin, because you would thereby win \$1m. More generally, it seems that there is a reason of the wrong kind for an attitude just when that attitude would be valuable in some way.<sup>29</sup>

This simple observation suggests that we might give an account of the wrong kind of reasons in terms of value. We might claim that:

---

<sup>29</sup> It might be thought that the wrong kind of reasons have to do specifically with the *instrumental*, or at least *non-final*, value of an attitude. However, it will be simpler to ignore this complication.

(Value-Based Theory) For  $p$  to be a reason of the wrong kind for attitude A is for  $p$  to be a respect in which it is good to have attitude A.

Given the Value-Based Theory, an explanation of the Wrong Reason Pattern seems to fall into place. When there is a reason of the wrong kind for attitude A, there is some respect in which it is good to have attitude A. And when attitude B facilitates attitude A, there is also a respect in which it is good to have attitude B – attitude B facilitates attitude A, which it is good to have. So by the Value-Based Theory, the fact that attitude B facilitates attitude A is a reason of the wrong kind for attitude B.

To take an example, when there is the wrong kind of reason to intend to drink the toxin, there is a respect in which it is good to intend to drink the toxin. And so if believing that you ought to drink the toxin facilitates this intention, there is also a respect in which it is good to believe that you ought to drink the toxin – this belief facilitates the intention to drink the toxin. So by the Value-Based Theory, the fact that believing that you ought to drink the toxin facilitates intending to drink the toxin is a reason of the wrong kind to believe that you ought to drink the toxin.

The Value-Based Theory thus seems to offer a straightforward explanation of the Wrong Reason Pattern. However, in the present context, there is an immediate worry about this explanation. While there is more than one reason to be interested in the distinction between the right and the wrong kind of reasons, the WKR defenders we are concerned with here make this distinction, at least in part, to solve the WKR problem for FA accounts.<sup>30</sup> But value-based theories of reasons seem to be in clear tension with FA accounts. It looks circular to hold both that the valuable is what there is reason to value, and that reasons to value are to be understood in terms of what is valuable.

---

<sup>30</sup> WKR defenders who do not wish to defend FA accounts do not face the worry I go on to discuss. This is an important restriction on the scope of my argument.

However, there is a way for WKR defenders to break out of this circle. Proponents of FA accounts must certainly deny that *all* reasons are to be understood in terms of values. But the Value-Based theory is an account of only the *wrong* kind of reasons. So long as there is an independent and non-value-based account of the right kind of reasons, WKR defenders can thus accept the Value-Based Theory and an FA account, and so explain the Wrong Reason Pattern.

However, combining an FA account with the Value-Based theory leads us back to Reductionism. The Value-Based theory says that for  $p$  to be a wrong kind of reason for attitude  $A$  is for  $p$  to be a respect in which it is good to have attitude  $A$ . FA accounts will say that for  $p$  to be a respect in which a thing is good is for  $p$  to be a reason of the right kind to favour that thing. So given an FA account, the Value-Based theory is simply another way of stating the thesis of Reductionism:

(Reductionism)      For  $p$  to be a reason of the wrong kind for attitude  $A$  is for  $p$  to be a reason of the right kind to want attitude  $A$ .

### 4.3. Reductionism

We have now seen that both of the WKR defender's strategies for explaining the Wrong Reason Pattern involve a commitment to Reductionism. To see whether these strategies are successful, we must thus consider the merits of this thesis.

The difference between Reductionism and WKR skepticism may seem slight. WKR skeptics hold that incentives for attitudes are reasons to want (and bring about) those attitudes, but not reasons for those attitudes. Reductionists hold that incentives for attitudes are reasons for those attitudes, but that what it is for an incentive to be a reason for an attitude

is for it to be a reason of the right kind to want that attitude. So whereas WKR skeptics appeal to reasons to want attitudes to explain why we *mistake* incentives for attitudes for reasons for those attitudes, Reductionists appeal to such reasons to *analyse* the wrong kind of reasons. The central challenge for Reductionists is thus to show that the right kind of reasons to want an attitude also count as, or ‘add up to’, reasons for that attitude. To put it another way, Reductionists must show that their analysis entails that the wrong kind of reasons for an attitude really are reasons for that attitude. I shall argue that Reductionists face serious difficulties meeting this challenge.

We can start to see the problem by contrasting the Reductionist account of the wrong kind of reasons with the accounts of the right kind of reasons discussed above. Each of these accounts works by taking for granted the relation of being a reason for, and then pointing to a further mark or feature that distinguishes the right kind of reasons (e.g. that they are ‘object-given’, or bear on correctness, or are shared by everyone engaged in a certain activity). In this way, the right kind of reasons are explained in the way it is natural to explain what it is to be an isosceles triangle, or a snub nose. We say what an isosceles triangle is by saying that it is a triangle with a certain feature – namely, that two of its sides are of the same length. And we say what a snub nose is by saying that it is a nose with a certain feature – namely, concavity.<sup>31</sup> Similarly, the accounts we have considered say what a right kind of reason for an attitude is by saying that it is a reason for that attitude with a certain further feature. Thus we may say that each of these accounts offer a *conjunctive analysis* of the right kind of reasons.

Conjunctive analyses have important advantages. In particular, such analyses trivially meet the sort of challenge faced by Reductionism. It is a trivial consequence of the accounts above that the right kind of reasons for an attitude are reasons for that attitude, that isosceles triangles are triangles, and that snub noses are noses.

---

<sup>31</sup> I take the latter example from Anton Ford, who takes it from Aristotle. See ‘Action and Generality’, in *Essays on Anscombe’s Intention*, ed. Anton Ford, Jennifer Hornsby, and Frederick Stoutland (Oxford: Oxford University Press, 2011), 76-104.

Reductionism is not a conjunctive analysis of the wrong kind of reasons. It does not say that the wrong kind of reasons for an attitude are those reasons for that attitude which have a certain feature. And so it is not a trivial consequence of Reductionism that the wrong kind of reasons for an attitude are reasons for that attitude. Reductionists thus have further work to do to show that their account has this consequence.

There are at least three ways in which Reductionists might try to meet this challenge. First, Reductionists might simply revise their account so that it becomes a conjunctive analysis. They might claim that for  $p$  to be a wrong kind of reason for attitude A is for  $p$  to be a reason for attitude A *and* a right kind of reason to want attitude A. However, if Reductionists revise their account in this way, their view no longer explains the Wrong Reason Pattern. To explain the Wrong Reason Pattern, Reductionists would now need to show that when there is the wrong kind of reason for attitude A, the fact that attitude B facilitates attitude A is both a reason for attitude B and a reason of the right kind to want attitude B. The Desire Pattern ensures the second of these conditions is met, but it does not guarantee that the first is.

Second, Reductionists might look for a non-trivial way of showing that the right kind of reasons to want an attitude are also reasons for that attitude. To see how this might work, note that there can be different levels of analysis.<sup>32</sup> We can analyze isosceles triangles as triangles with two equal sides, but we can also say – going a bit deeper – that isosceles triangles are three-sided closed figures with two equal sides. While this latter analysis is intuitively deeper, it does not trivially entail that isosceles triangles are triangles. But when we add that triangles are three-sided closed figures, we can see that it entails it nonetheless. Similarly, it might be that when we combine Reductionism with a general account of what it

---

<sup>32</sup> See Jeffrey King, 'What Is A Philosophical Analysis?' *Philosophical Studies* 90 (1998): 155-79, 167.

is to be a reason for an attitude, it will become clear that Reductionism does, after all, entail that the wrong kind of reasons for an attitude are reasons for that attitude.

It is difficult, however, to see how this might work. The suggestion is that Reductionism might implicitly include an analysis of what it is to be a reason for an attitude, just as the second, deeper, analysis of what it is to be an isosceles triangle implicitly includes an analysis of what it is to be a triangle. But there is nothing in Reductionism which looks as if it could play that role. The Reductionist's analysis involves two constituent properties – that of being a reason to want an attitude, and whatever further property distinguishes the right kind of reasons. But on pain of circularity, we cannot say that what it is for something to be a reason for an attitude is for it to be a reason to want that attitude. And nor can the further property serve as such an analysis. This further property is precisely intended to pick out only the *right* kind of reasons – it cannot be the sort of thing which will characterize *all* reasons for attitudes.

Third, Reductionists might suggest that we take the right kind of reasons to be prior to reasons in general, and offer a disjunctive account of the latter in terms of the former. On such an account, what it is for *p* to be a reason for some response is for *p* to be *either* a right kind of reason for that response *or* a right kind of reason to desire that response. Such an account would trivially imply that the right kind of reasons to desire an attitude are reasons for that attitude. However, while this may be the most promising of the Reductionist's options, it nonetheless faces at least two significant problems. First, since it requires us to give up the natural thought that the right kind of reasons should be given a conjunctive analysis, it is incompatible with at least some of the familiar accounts of the right kind of reasons discussed above, all of which are naturally formulated as conjunctive analyses.<sup>33</sup>

Second, the suggestion faces a general worry about disjunctive analyses. If it turns out that

---

<sup>33</sup> Some of these accounts – such as the 'Brentano-style' account and D'Arms and Jacobson's account – may be reformulable as non-conjunctive analyses. It is harder to see how to reformulate the object/state theory and Schroeder's account as non-conjunctive analyses.

what it is for something to be F is for it to be either G or H, then it looks as if it has turned out that the Fs do not have anything distinctive in common – as it turned out that pieces of jade do not have anything distinctive in common.<sup>34</sup> It is a cost of an account if it implies that a category of things which appear to have something distinctive in common – such as reasons – turn out not to do so.

It thus seems to me that none of the three ways by which Reductionists might try to show that the right kind of reasons to want an attitude are also reasons for that attitude is satisfactory. If this is so, WKR defenders should not accept Reductionism.

## **5. Conclusion**

Both of the WKR defender's strategies for explaining the Wrong Reason Pattern commit the WKR defender to Reductionism. But Reductionism, I have argued, should be rejected. WKR defenders are thus left without a satisfactory explanation of the Wrong Reason Pattern. By contrast, WKR skeptics offer a straightforward explanation of why, if we mistake reasons to want and bring about an attitude for reasons for that attitude, we will also take the Wrong Reason Pattern to hold. I take this to be an important advantage of WKR skepticism, and an important problem for WKR defenders. However, my argument faces two questions.

First, like all arguments of this form, my argument is open to the charge that I have overlooked promising alternative strategies by which WKR defenders might try to explain the Wrong Reason Pattern. To this charge, there is nothing very definitive I can say. I do not know how to argue that WKR defenders have no promising options beyond the ones I have discussed. All I can say is that I cannot think of any. It is of course possible that someone will

---

<sup>34</sup> Schroeder, *Slaves of the Passions*, 69.

come up with a better alternative.<sup>35</sup>

Second, it might be wondered where my argument leaves us. Suppose that it is granted that WKR skepticism offer the best explanation of the Wrong Reason Pattern. This would not show that WKR skepticism is true. Rather, it would provide one consideration in its favour. But it may be, as several philosophers have recently argued, that there are also important considerations counting against WKR skepticism. If these philosophers are right, then perhaps we should reject WKR skepticism, its capacity to explain the Wrong Reason Pattern notwithstanding.

I can say more about this concern. While several philosophers have offered objections to WKR skepticism, I do not think any of these objections succeed. By way of a coda, I shall argue that WKR skeptics can make entirely satisfactory responses to three central objections that have been raised against them. If these responses succeed, then it seems to me that the WKR skeptic's straightforward explanation of the appearance of the Wrong Reason Pattern does tip the balance of considerations in favour of WKR skepticism.

Two points will be central to my case. First, we can often see how to respond to objections to WKR skepticism by considering why parallel objections do not arise for the view that incentives for headaches and perceptual experiences do not provide reasons for these things. Second, while WKR skeptics can allow that in many cases incentives for attitudes are reasons both to want and to bring about those attitudes, WKR skeptics should take the former to be primary.

WKR skepticism is a kind of error-theory. It holds that people mistakenly think that incentives for attitudes are reasons for those attitudes because they confuse reasons for

---

<sup>35</sup> One possibility is to appeal to a *desire-based* theory of the wrong kind of reasons. Such a theory would hold that there is a wrong kind of reason for an attitude just when having that attitude facilitates something you desire. Since such views are structurally similar to value-based theories – both explain reasons for attitudes by pointing to something facilitated by attitudes – such views are of the right form to explain the Wrong Reason Pattern. However, desire-based theories of the wrong kind of reasons face many of the familiar problems faced by desire-based theories of reasons for action (see, e.g. Parfit, *On What Matters*, chaps. 3 and 4). For example, even if you do not care about being tortured, it seems like the demon's threat gives you a reason to admire him.

attitudes with reasons to want or bring about an attitude. One way to argue against WKR skepticism is thus to argue that there are cases in which there is an incentive for an attitude but no reason to want or to bring about that attitude. I shall begin by considering several putative such cases. I shall argue that while there are several kinds of case in which incentives are not reasons to bring about the relevant attitudes, in all such cases there are still reasons to want to those attitudes. The moral of these examples is thus that WKR skeptics should take reasons to want attitudes as primary.

First, there are cases in which there is simply nothing you can do to bring about the relevant attitude. If the prize is for believing, *right now*, that the Moon is made of green cheese, then unless you happen to have on you a very fast working “belief pill”, you simply will not be able to do anything to bring about this belief. And so if we think that ‘reason’ implies ‘can’, you will not have any reason to bring about the belief that the Moon is made of green cheese. Nonetheless, you have a reason to want to believe that the Moon is made of green cheese.

Second, there are cases in which there is no need to bring about the relevant attitude. As Andrew Reisner observes, the incentive might be offered for a belief that you already hold, or for which you have been presented with conclusive evidence.<sup>36</sup> In these cases, bringing about this belief is unnecessary and so not plausibly something you have reason to do. Nonetheless, you have a reason to want this belief.

Third, there are cases in which the incentive holds on the condition that you do not bring about the relevant attitude.<sup>37</sup> This is the structure of Kavka’s original toxin puzzle – as Kavka describes things, the billionaire will reward you only so long as you intend to drink the toxin *without* having hired a hypnotist, or created extra incentives for drinking the toxin, or done any of the other things which might help to bring about this intention. In cases of this

---

<sup>36</sup> Reisner, ‘The Possibility of Pragmatic Reasons’, 270-1.

<sup>37</sup> Reisner (*ibid.*, 271) calls these cases of ‘blocked ascent’.

sort, incentives are not reasons to bring about the relevant attitudes. Nonetheless, they are reasons to want these attitudes.

However, it might be thought that we can also describe cases in which incentives are not reasons to want the relevant attitudes. As Reisner points out, there is nothing to stop evil demons, or eccentric billionaires, offering incentives for attitudes which are conditional on your lacking the desire for that attitude.<sup>38</sup> For instance, the evil demon might threaten to kill you unless you admire him without also wanting to admire him. Reisner claims that in cases of this sort, incentives are not reasons to want the relevant attitude.

It is not clear that this is right. The further conditions in these cases certainly give you an *incentive* not to desire the relevant attitudes. But that is compatible with there also being reasons to want these attitudes.<sup>39</sup> So more needs to be said to defend the claim that the incentives in these cases are not reasons to want the relevant attitude.

There is more that can be said. If the incentives in these cases are reasons to want the relevant attitudes, they are reasons of a rather strange sort. For notice that if you form the desire to admire the demon in response to the demon's threat, then the incentive for admiring the demon disappears – the demon will kill you anyway. So if the incentive for admiring the demon is a reason to want to admire the demon, it is a reason which cannot survive being responded to.

It is a matter of dispute whether there can be reasons of this sort. Many philosophers accept an “internalist” condition on reasons, according to which a consideration is a reason to A only if it is possible to A for that reason. This condition rules out reasons of the above sort. However, many other philosophers reject this condition, and hold that there can be reasons

---

<sup>38</sup> Ibid., 271.

<sup>39</sup> Richard Chappell has independently made this point. See <http://www.philosophyetc.net/2009/12/who-has-intuitions-about-whether-there.html>.

which cannot be responded to.<sup>40</sup> The above example shows that WKR skeptics are committed to joining this latter camp.

This may be thought to count against WKR skepticism. But in fact examples of this sort show that *all* proponents of FA accounts – WKR skeptics and defenders alike – are committed to accepting reasons of this sort. In the example imagined, it would clearly be good to admire the demon, since doing so would save your life. FA accounts are thus committed to thinking that there are reasons to want to admire the demon, even though responding to this reason would make it disappear. So the claim WKR skeptics need to make about Reisner's example is one that FA accounts need to make anyway.

A second objection to WKR skepticism claims that there is a tension between the two claims WKR skeptics make – that incentives are reasons to want and to bring about attitudes, but not reasons for those attitudes. Several philosophers worry that the latter, negative, claim undermines the former, positive, claim. Thus Rabinowicz and Rønnow-Rasmussen write, 'To be sure... we have reasons to want to have such attitudes and to try to have them, but this is only because we have reasons to have them, in the first place'.<sup>41</sup>

There are at least three ways to understand this objection. A first version starts with the premise that there is reason to want or bring about an attitude only if there is a reason for that attitude. It follows immediately from this premise that WKR skepticism is false. However, we should not accept this premise without argument. After all, the WKR skeptic's central claim is precisely that there are cases – the WKR scenarios – in which there are reasons to want or bring about attitudes there is no reason for. To simply assert otherwise is to beg the question.

---

<sup>40</sup> See Elijah Millgram, 'Williams' Argument Against External Reasons', *Nous* 12 (1996): 345-64; Schroeder, *Slaves of the Passions*, 165-7; and Julia Markovits, 'Internalism and the Motivating Intuition', in *New Waves in Metaethics*, ed. Michael Brady (Basingstoke: Palgrave Macmillan, 2010) (and see references therein).

<sup>41</sup> 'The Strike of the Demon', 413. Cf. Olson, 'Buck-Passing and the Wrong Kind of Reasons', 297, Louise, 'Correct Responses', 350, Markovits, 'Internalism and the Motivating Intuition', 155.

A second version of the objection aims to defend the crucial premise by appealing to a generalization of it – that there is reason to want or bring it about that you F only if there is reason for you to F. However, this generalization is certainly false. Suppose, for instance, that you are offered a large prize for having a headache, or an experience of redness. You would then have reason to want or to bring about a headache or an experience of redness, although there are no reasons for headaches or experiences of redness. Not everything there is reason to want is something that there is reason for.

A third version of the objection is more promising. Reisner suggests that WKR defenders can offer a straightforward explanation of the reasons to bring about attitudes for which there is an incentive, which is not available to WKR skeptics.<sup>42</sup> The explanation appeals to the transmission of reasons across facilitative connections. Reisner's thought is that WKR defenders can say that there is reason to bring about such an attitude because bringing about such an attitude is a means to having an attitude which there is reason for. However, this explanation requires there to be reasons for such attitudes, and so is not available to WKR skeptics.<sup>43</sup>

Reisner is right that WKR skeptics cannot explain reasons to bring about these attitudes in this way. But an equally plausible explanation is available to them. For as a first pass, it is highly plausible that there is reason to bring about states that there is reason to want. Something along these lines is needed to explain, for instance, why there can be reasons to bring about headaches and perceptual experiences, among other things there can be reason to want, but which are not themselves subject to reasons. And this principle can equally explain why there are reasons to bring about attitudes for which there are

---

<sup>42</sup> Reisner, 'The Possibility of Pragmatic Reasons', 269-70.

<sup>43</sup> There is room to challenge the transmission pattern which Reisner relies on here. Reisner's pattern requires there to be reasons to do things that facilitate *attitudes* there is reason to have. I think the arguments against the Mixed Pattern in section two also count against this pattern. Nonetheless, I will grant Reisner's pattern for present purposes.

incentives.<sup>44</sup> So this version of the objection also fails to show that the WKR skeptic's negative claim undermines the positive claim.

There is a final objection to WKR skepticism to consider. Rabinowicz and Rønnow-Rasmussen end their discussion of WKR skepticism by claiming that WKR skeptics have no real advantage over WKR defenders, with respect to the WKR problem. This is because WKR skeptics will still 'need to clarify, without taking the notion of 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'. Rabinowicz and Rønnow-Rasmussen suggest that this project will in effect require a further solution to the WKR problem.<sup>45</sup>

This objection can be understood in two ways. In saying that WKR skeptics must say 'what makes something a reason', Rabinowicz and Rønnow-Rasmussen could either be asking for an analysis of what it is to be a reason to want, or they could be asking for a substantive account of the conditions under which there are reasons to want.

However, WKR skeptics should accept neither of these demands. As is often pointed out, it is a virtue of FA accounts that they are neutral on the substantive question of what things are of value – e.g. between hedonism, desire-based theories, and the various objective-list theories. However, it is clear that this neutrality requires FA accounts to be neutral on the substantive question of when exactly there are reasons to value. The only substantive claims WKR skeptics need to make are those that define the view – that incentives for attitudes are not reasons for those attitudes but are reasons to want, and in many cases bring about, those attitudes. These claims are compatible with a wide-range of substantive views of the conditions under which there is reason to want something. Nor need WKR skeptics accept a

---

<sup>44</sup> Of course, this response leaves reasons to want such attitudes unexplained. But the WKR defender has no advantage here. Both WKR skeptics and defenders can explain reasons to want these attitudes by appealing to the Desire Pattern.

<sup>45</sup> 'The Strike of the Demon', 413-4. Cf. Louise, 'Correct Responses', 350.

demand to provide an analysis of reasons to want. For it is equally a virtue of FA accounts that they are compatible with a wide-range of views of what it is for something to be a reason, as well as with the view that reasons are unanalyzable. And again a wide-range of such views are compatible with the WKR skeptic's claims.

I conclude that none of the central objections to WKR skepticism in the recent literature show WKR skepticism to be implausible. So if the central arguments of the paper succeed, WKR skeptics really do have a significant advantage over WKR defenders.