

## THE RELEVANCE OF THE WRONG KIND OF REASONS

ULRIKE HEUER

UNIVERSITY OF LEEDS

*“Let us weigh the gain and the loss involved in wagering that God exists. Let us estimate these two probabilities; if you win, you win all; if you lose, you lose nothing. Wager then, without hesitation, that He does exist.” (Blaise Pascal, *Pensées*)*

That you win doesn't seem the right kind of reason for believing that God exists, even if Pascal is right that you stand to lose nothing, but you may gain a lot.<sup>1</sup> It seems to be a reason of the wrong kind. There is currently a wide-ranging philosophical discussion of two kinds of reasons for attitudes such as belief, which are sometimes called the right and wrong kinds of reasons for those attitudes. The question is what the distinction shows about the nature of the attitudes, and about reasons and normativity in general. The distinction is deemed to apply to reasons for different kinds of attitudes such as beliefs and intentions, as well as so-called pro-attitudes, e.g. admiration or desire.

Wlodek Rabinowicz's and Toni Rønnow-Rasmussen's paper "The Strike of the Demon" (2004) forged the shape and structure of the current discussion. Their focus is on the fitting-attitude theory of value (henceforth: FA theory), a version of which is (as they see it) the so-called 'buck-passing account'<sup>2</sup>, the view that the value of an object consists in the fact that it has properties other than that of being valuable which provide reasons for pro-attitudes. The main problem for this approach is that there are some reasons for having those attitudes which are not provided by properties that make their object valuable. Those are the wrong kind of reasons in the sense that FA theory must be able to distinguish them from those properties that figure in the analysis of value.

But the discussion of the wrong kind of reasons problem (for short: WKR problem) has spread far beyond FA theory and buck-passing accounts: various philosophers<sup>3</sup> hope to provide a unified account of central issues concerning practical and theoretical normativity by giving an explanation of why certain reasons are of the wrong kind for forming an attitude.

They perceive what appears to be a unified phenomenon, and suspect that the explanation of why some reasons are of the wrong kind and others of the right kind for having an attitude are relevantly similar across different attitudes, and may

---

<sup>1</sup> Note, however, that Pascal does not enjoin us to *believe*, but to *wager* that God exists.

<sup>2</sup> Scanlon (1998); for a critical discussions of the buck-passing account, see Heuer (2006 and 2011) and Väyrynen (2006).

<sup>3</sup> E.g. Hieronymi (2005), Shah (2008), McHugh (2012), Way (2012), Schroeder (2012).

reveal constitutive normative constraints<sup>4</sup> that the attitudes in question are subject to. If the explanation of the right kind/wrong kind of reasons distinction is basically the same, or has the same general form, for theoretical attitudes such as belief, and practical attitudes like intention, and both rely on the normative constraints that are constitutive for the attitudes, then we may be able to provide a unified account of theoretical and practical normativity. Thus, explaining why certain reasons are of the wrong kind would furnish a heuristic for discovering important features of practical and theoretical normativity alike.

In this paper I want to show why this ambition is misguided. I will argue for two claims: (1) we should sharply distinguish the wrong kind of reasons problem as it arises for FA theories from any other problem that comes under the same name, and (2) the WKR problem outside of FA theory doesn't have a very clear shape (if indeed there is such a problem at all). In particular, there is no similarity between reasons to believe and reasons to intend in this regard, and therefore no hope for a unified explanation of the alleged phenomenon.

### 1. The WKR problem in FA theory

FA theory which led to formulating the WKR problem is concerned with reducing the evaluative to the deontic. That something is of value means (semantic reduction) or consists in (metaphysical reduction) the fact that it is fitting to have certain attitudes towards it. When are attitudes of the relevant kind (roughly, pro-attitudes) fitting or appropriate? If the answer were that they are appropriate when their object is of value, the reduction would have failed. If having value reduces to the fittingness of those attitudes, then what makes the attitudes fitting cannot be that the object has value: the fitting attitude cannot explain itself.

The so-called buck-passing account of value avoids this circle: According to it, that an object has value consists in its having *other* properties that make certain attitudes towards the object appropriate (or: that provide reasons for having those attitudes). Being of value reduces to having those other properties that make the attitude fitting. Understood in this way, the buck-passing account is an improved version of FA theory.<sup>5</sup>

The WKR problem then arises because, intuitively, there are reasons for attitudes that are unrelated to the value of their objects. Evil-demon examples have been used to illustrate the point: that an evil demon threatens to punish you if you don't admire a worthless object may be a reason for admiring it, but not a reason that is provided by a property that makes the object valuable.

---

<sup>4</sup> Or alternatively, the explanation may be teleological appealing to the 'aim' of the attitude (cf. McHugh 2011, 2012; for a critical discussion see Owens 2003).

<sup>5</sup> McHugh and Way (forthcoming) show that the move by which the buck-passer avoids the circle is available to FA theorists as well, even if they do not endorse buck-passing. They suggest that we should distinguish between two claims: (1) that value consists in have other properties that make certain attitudes fitting, and (2) that 'making fitting' is the same as providing reasons. McHugh and Way argue that the FA theorist should accept the first claim (thereby avoiding the circle) but not the second. Buck-passing views endorse both.

Unless the FA theorist can identify the right kind of reasons for having the relevant attitudes, the analysis fails: it doesn't provide an account of value. The wrong kind of reasons problem is the problem that there seem to be counterexamples to the proposed analysis of value. They prove it wrong – unless it is possible to distinguish between the right and the wrong kind of reasons, and to reformulate the account in terms of the right reasons.

FA theory needs to explain which reasons are the right ones while avoiding circularity. The answer should *not* be that the right kind of reasons for an attitude are those that are provided by the properties that make their object valuable. The explanation to be avoided looks somewhat like this:

**[Value]** The value of an object, *O*, consists in the fact that *O* has other properties that provide the right kind of reasons for certain favorable attitudes towards *O*.

**[RK Reasons]** Reasons of the right kind are those that are provided by those properties that make *O* valuable.

The buck-passing account avoids the simpler circle of explaining the right kind of reasons in terms of the object's value by emphasizing that the reasons are provided by properties other than that of being valuable, but it doesn't help with this one.<sup>6</sup>

This is the WKR problem for FA theory. Proponents of FA theory have provided ingenious answers to the circularity worry.<sup>7</sup> My concern here is not with the possibility of solving this problem, but with distinguishing it from a different problem with which it is sometimes conflated. There is a wider use of the right/wrong kind of reasons terminology which has gained currency independently of FA theory. It is used to draw a distinction between reasons for attitudes in general. (Some even think that there is such a distinction with regard to all reasons, including reasons for action.<sup>8</sup>)

Let's call those who use the wrong/right kind of reasons distinction outside of FA theory *Reasons Theorists*. Their hope is to provide a unified account of theoretical and practical normativity by focusing on the WK/RK reasons distinction as a shared feature in particular of reasons to believe and reasons to intend.

---

<sup>6</sup> Rabinowicz and Rønnow-Rasmussen (2006) explore the possibility of embracing the circle. While circular explanations can be helpful, this particular circle seems explanatorily void to me. The only thing we learn from it is that values provide reasons. But that claim can be couched in a much less problematic form: it needn't be put as a metaphysical reduction. And it is this, ambitious, feature of the account that creates the problems.

<sup>7</sup> Cf. Rabinowicz & Rønnow-Rasmussen (2004, 2006), Olson (2004), Lang (2008), Way (2012).

<sup>8</sup> Schroeder (2010). There is indeed such a distinction within FA theory: there may be actions which could count as pro-responses to valuable objects in the relevant sense. E.g., there may be a reason for a pro-response towards an object, such as protecting it, because the object has final value. If so, there can be reasons of the wrong kind for this response, 'protecting', i.e. reasons that are not provided by those properties that make the object valuable. I am grateful to Wlodek Rabinowicz for alerting me to this possibility. I will come back to it in section 2, fn 19. But, *pace* Schroeder (2010), there are no wrong kind reasons for actions beyond FA theory (Heuer 2011).

Specifying the two claims I will argue for in the light of this distinction, they are, first, that the Reasons Theorist must mean something different from what FA theorists mean by 'the right/wrong kind of reasons' (section 2), and secondly, that there is no WKR problem, remotely like the one in FA theory, which arises for reasons for attitudes in general. In particular, it is not a unifying feature of reasons for attitudes in general that we can draw a wrong/right kind of reasons distinction. Whilst reasons to believe may appear to support the Reasons Theorist's view, reasons to intend do not: there cannot be a unifying account of practical and theoretical normativity along these lines. (Section 3)

## 2. Cutting loose: The alleged WKR problem beyond FA theory

While it is quite clear what is meant by the right/wrong kind of reasons in FA theory, it is pretty unclear beyond. In FA theory the wrong kind of reasons are simply putative counterexamples to the theory. But what do those who hope to use the right/wrong kind of reasons distinction beyond FA theory mean by it?

A core example where the right/wrong kind of reasons distinction is deemed to have application – outside FA theory – is belief. Traditionally, FA theory does not make reference to belief, since it is not a pro-attitude.<sup>9</sup> But might there nonetheless be a similar distinction of right/wrong kind reasons for belief? After all, with regard to beliefs we can distinguish between reasons to believe that  $p$  because there is evidence for  $p$ , and reasons to believe that  $p$ , because having the belief would be good or beneficial in some way (e.g. Pascal's wager, or a reason for believing  $p$  because it would make one happy if  $p$ ). Evidence for the truth of  $p$  is a kind of reason for believing  $p$  that we can follow, and perhaps even follow automatically, when considering whether  $p$ . In this regard it contrasts with other kinds of reasons for having beliefs. That it would make me happy to believe that the EU referendum (in the UK) was in favour of remaining may be a reason for having that belief, but I cannot form it for that reason. It would also not be rational to form the belief for that reason, and it wouldn't be a failure of rationality to not respond to this reason at all. Furthermore, *one cannot reason to the belief that  $p$  (or that not- $p$ ) via the reason that believing  $p$  (or: not- $p$ ) causes happiness or anxiety*. So those are (perhaps) reasons of the wrong kind for believing  $p$ .

Are they tracing the same distinction that the FA theorist needs to draw with regard to attitudes such as admiration, desire, or preference? At first blush, the similarity seems striking. After all, as with wrong kind reasons in FA theory, the reason is provided by the value of having the attitude and in both cases, it seems that it isn't possible to form the attitude for this kind of reason – at least not directly. E.g. it is not possible to admire a worthless object because, unless you do, an evil demon will torture you. Having the attitude would be good in this case, but you cannot form it for that reason – at least not directly.

---

<sup>9</sup> But see McHugh (2014).

Mark Schroeder<sup>10</sup> has suggested a set of four “earmarks” to capture the characteristic features of the right/wrong kind of reasons distinction generally (i.e. within FA theory and beyond). Formulated with regard to belief, the earmarks are:

1. **“Asymmetry of motivation”**: “it is easier” to believe that  $p$  on the basis of evidence for the truth of  $p$  than to believe that  $p$  as, say, the result of wishful thinking.
2. The wrong kind of reasons are reasons that do not contribute to the **rationality** of believing  $p$ .
3. They also don’t bear on the **correctness** of believing  $p$ .
4. They have a distinctive **“flavour”**.

We should bracket (3) and (4). ‘Correctness’ in (3) appeals to the claim that a belief is deemed correct if it is true. But if so, there can be any number of reasons of the right kind that do not bear on the belief’s correctness: all those where the evidence is misleading.<sup>11</sup> A person who forms a belief on the basis of misleading evidence may well be rational even if her belief is false. That your friend’s phone number appears on your phone’s display is a reason to believe that he is calling you and you are rational in forming the belief, even if in fact his phone has been stolen, and the caller is the thief. Thus (2) and (3) pull in different directions. In addition, it is not clear what ‘correctness’ comes to with regard to attitudes other than belief. There is no standard of correctness comparable to truth with regard to intentions.<sup>12</sup> Hence this isn’t a feature which is shared by different kinds of attitudes.

(4), as far as I can make out, doesn’t add anything. This leaves us with (1) and (2).

(1) is, I think, misleadingly put. It may be very difficult to believe that, say, one’s best friend committed a heinous crime, but the evidence that she did is not therefore of the wrong kind. Whether or not it is difficult to form the attitude is irrelevant. I suspect that (1) is inspired by a different kind of concern anyway: by the observation that we can follow only certain kinds of reasons directly when we form a belief (or reason towards a belief). We should then replace (1) with the stricter (1\*): **we cannot reason to the belief that  $p$  via a certain kind of reason for believing that  $p$ .**<sup>13</sup>

---

<sup>10</sup> Schroeder (2012), 458ff.

<sup>11</sup> The referees urged me to clarify: doesn’t misleading evidence bear on a belief’s correctness in the sense that there must be some connection to the belief’s truth for it to be evidence at all? I’m not sure of this. My evidence can be misleading in the sense that if I had reasoned correctly (if I had been aware of defeaters, for instance), I would not have formed the false belief. It can also be misleading in the sense that even if I reason correctly, I end up with a false belief. The connection with truth in the second case is moot. But be that as it may, the sense of ‘bear on’ that I’m concerned with is that following misleading evidence does not contribute to forming a true belief.

<sup>12</sup> Schroeder admits as much, and reserves (3) for beliefs only. If so, it will not help with finding a unified account of wrong kind reasons for attitudes in general anyway. For a diverging view, attempting to formulate a standard of correctness for intentions, see Shah (2008) and McHugh (2012).

<sup>13</sup> Schroeder identifies putative WK reasons as ‘state-given’ reasons. Since I don’t find the distinction of state- and object-given reasons helpful, I will stick for the time being with the more concrete example of wishful thinking-type reasons. I will come back to the question whether we can identify the relevant kind of reasons in a more general way.

In addition, we may want to accept (2): purported wrong kind reasons do not contribute to the rationality of forming the attitude. This is the feature I relied on when I described reasons in wishful thinking cases as putatively of the wrong kind.<sup>14</sup>

If (1\*) and (2) identify wrong kind reasons, the FA theorist might be able to utilize those earmarks to solve her WKR problem: to give a non-circular account of the counterexamples to the view. The suggestion would be that a putative reason is of the wrong kind if and only if we cannot reason from it to the attitude for which it is a reason, and responding to the reason would not contribute to the rationality of the attitude in question (or: not responding to it does not prove the person irrational, even if she is aware of the reason, and believes that it is conclusive). If so this could be our candidate explanation of the wrong/right kind of reasons distinction, cutting across FA theory, and Reasons Theory.

Let us test this suggestion by considering how it might be applied in FA theory. There is relative agreement that evidence for the truth of a proposition,  $p$ , is a reason to believe that  $p$ . One kind of evidence is testimony. Testimony and other kinds of evidence are reasons of the right kind for forming beliefs, if anything is. I can rationally form the belief that it is now one o'clock on the basis of your testimony.

Compare now reasons to believe with reasons for different kinds of attitudes, the ones that FA theory is in fact concerned with. 'Admiration' is one of those. Reasons to admire someone or something may be evidence- or testimony-based. Perhaps I admire a medical researcher because a friend who (unlike me) understands medical research has assured me that the researcher's work is brilliant. Or I admire Einstein because his work is deemed by everyone in the profession as ground-breaking and setting the agenda for modern physics. It seems that my admiration would be appropriate in these cases. But the reason for it – that my friend told me, or that physicists agree that the researcher, or Einstein, is brilliant – is not one of the features that make it the case that she or he is. Their brilliance does not consist (even in part) in my friend's or their peers' testimony.

For the FA theorist, evidence-based reasons are of the wrong kind for admiration (assuming that testimony is a kind of evidence). It is a counterexample to the basic FA-formula: "The researcher,  $A$ , is admirable if and only if  $A$  has other properties that make admiration towards her appropriate." Whatever properties the medical researcher may have that make her admirable, my friend's testimony is not among

---

<sup>14</sup> Making the connection with rationality requires understanding evidence as subjective evidence: considerations that a person takes to be evidence. It is, at any rate not clear, whether disregarding evidence for  $p$  that a person does not regard as evidence for  $p$  proves a rational failure. 'Right reasons' would then be all those reasons that make an attitude rational, even if they include misleading evidence, or are shaped by a failure to take in relevant information, as long as this isn't a failure of rationality. Hieronymi's astrology example is an ambiguous one here: as she sees it, a person who believes that the stars influence our destiny, and reads in her daily horoscope that she will have a wonderful day today, might believe this for a right kind reason. Right kind reasons, Hieronymi (2005) urges, are not 'good reasons'. On the current approach, believing that the stars influence our destiny must either be a belief that one could rationally hold, or her understanding of right reasons does not endorse the two earmarks (but perhaps only 1\*). However, her general conclusion that right reasons needn't be good reasons holds either way.

them. But even though the testimony-based reason is of the wrong kind, it is not a reason that we cannot follow or reason from, or a reason that doesn't contribute to the rationality of the attitude.

There are evidence-based reasons for other pro-attitudes as well: I may desire a new computer because it received good reviews. For FA theory this is a wrong kind reason to desire the computer: the value of the computer does not consist in its getting good reviews. (If anything, it gets good reviews because it is a good computer – not the other way around.)

In addition, there are evidence-based wrong kind reasons which do not involve testimony: imagine a patient exhibiting changes in skin colour that are characteristic of liver disease. His doctor relies on this evidence in issuing a prescription for medication perhaps because she works under conditions where it is not possible to investigate the state of the liver itself more thoroughly. (There is no negligence involved.) According to the Reasons Theorist she responds to right kind reasons in forming the belief that the treatment is necessary: the yellowing of the skin is, let's assume, sufficient evidence of liver disease. How about FA theory though? The doctor's willingness to help by prescribing the medication seems an appropriate attitude: the observed change in skin colour is a reason for the attitude that the doctor can reason from; her attitude is rational. However, according to FA theory, the right kind reasons for the attitude (willingness to help) are those provided by the badness of the patient's state. But the skin colour isn't bad for the patient, and therefore not a right kind reason to want to treat him – it is only correlated with something that is: with the liver disease.

Evidence-based reasons, it seems, are throughout of the wrong kind for FA theory. But they are otherwise of the right kind. Hence FA theorists and Reasons Theorists track a different distinction when they use the right/wrong kind of reason terminology.

Let me briefly consider two objections to this argument:

*First objection:* My friend's testimony is not a reason to admire the researcher, you may object, but just a reason to *believe* that there is a reason to admire her. Therefore, evidence-based reasons are not reasons that the FA theorist is concerned with (on the assumption that she doesn't attempt to give an account of the fittingness of beliefs).

*Reply:* It is not clear to me why the testimony would not provide a reason for admiration, and not just for beliefs about reasons for admiration. Presumably, the objection simply takes for granted that evidence and testimony bear only on belief.<sup>15</sup> I doubt that this is so. But let's grant the move for the sake of the argument. Even so, it would not show that the argument does not go through: if I have a reason to *believe* that there is a reason to admire X, it would be (other things being equal<sup>16</sup>)

---

<sup>15</sup> Attitudes like admiration seem to incorporate beliefs about the admirability of their object. So insofar as reasons for admiration are reasons for the incorporated belief, they should be evidence-based, just as the standard reasons for belief generally are.

<sup>16</sup> Things are not equal if the reason is defeated, say.

rational for me to admire X. It would therefore be fitting to admire X because of the testimony. I could also reason from my belief that there is a reason to admire X to forming the attitude, admiration (to the degree that admiration can be the outcome of reasoning at all<sup>17</sup>). This should suffice to show that the reason is not of the wrong kind, except for FA theory where it is a counterexample to the analysis. In my final example the case is even clearer: the patient's skin colour is hardly just a reason to believe that willingness to help would be an appropriate attitude. It is a reason for the attitude itself. But it is, for FA theory, a wrong kind reason.

*Second objection:* The reason to admire is not that my friend tells me, but that to which she testifies: the brilliance of the research.<sup>18</sup> In the examples, my epistemic access to the relevant feature is via testimony. And similarly, the doctor's reason for being willing to help is her patient's illness, and not his skin colour which is only indicative of the illness.

*Reply:* This is, I think, wrong: evidence can be misleading without therefore failing to be evidence. Thus while my friend's testimony is evidence for the brilliance of the research, the research may in fact not be brilliant (perhaps unbeknownst to my friend, it was plagiarized). In that case (according to the objection), I wouldn't have a reason to admire the research on the basis of my friend's testimony. But it seems, intuitively, that I do have such a reason. At the very least, I have a reason to believe that there is a reason to admire the research. So it would be rational for me to admire it – and thus we're back with the first objection. The doctor's example shows even more clearly that evidence is defeasible: the case may be unusual in that the change in skin colour is caused by something other than a liver disease. But that does nothing to undermine that the doctor responds to a right kind reason when she is willing to help because of the change in skin colour, according to the Reasons Theorist, and to a wrong kind reason, according to the FA theorist.

I take it then that the argument is sound: the distinction between right and wrong kind reasons is not the same in FA theory and in Reasons Theory.<sup>19</sup> Many ordinary reasons that make forming an attitude appropriate are wrong kind reasons from the perspective of FA theory. But then the aim of FA theory is to provide an account (or an analysis) of *value* – not of the appropriateness of attitudes. So even if most reasons which justify pro-attitudes and make them appropriate were reasons of the

---

<sup>17</sup> If this is doubtful, it is not because in this case I argue from my belief that there is a reason to admire to admiring it, but rather because admiration, being an emotion, may not be typically the result of reasoning. Assuming that it is nonetheless what Scanlon (1998) has called a 'judgment-sensitive attitude', i.e. an attitude that is sensitive to judgments about reasons for and against it, the relevant sensitivity may not typically be realized by generating the attitude in a process of deliberate reasoning.

<sup>18</sup> I'm grateful to Mark Schroeder for this suggestion.

<sup>19</sup> There are other examples that might show this. Wlodek Rabinowicz has added three other kinds in an unpublished paper. To mention but one: in fn 8 I remarked that there is a WK/RK reasons distinction regarding *reasons to act* in FA theory, but it does not have a counterpart in Reasons Theory. In this case too, WK reasons in FA theory do not bear the two 'earmarks'. They are simply reasons to, say, protect an object which are not provided by the object's final value. E.g. an evil demon threatens me unless I protect something worthless. There is no problem with reasoning from those reasons, nor would be irrational to follow them.



wrong kind, the FA theorist needn't worry. That ordinary (demon-free) reasons turn out to be of the wrong kind is not a problem for FA theory. But what this does show is that the Reason Theorist and the FA theorist must be talking about a different distinction, since right kind reasons for Reasons Theory turn out to be wrong kind reasons for FA theory. It also shows that FA theory cannot resort to (1\*) and (2) in order to draw the distinction between the right and the wrong kind of reasons. At least some counterexamples to FA theory – i.e. wrong kind reasons within FA theory – will be right kind reasons according to those earmarks.

But even if FA theory and Reasons Theory are concerned with different distinctions, there is, at least with regard to beliefs, a relatively clear distinction between evidence-based reasons, and (e.g.) wishful thinking-type reasons.

Our preliminary result is that there are certain reasons for believing  $p$  where we cannot reason from those considerations to believing that  $p$ , and it wouldn't be rational to form a belief that  $p$  for those reasons, nor a failure of rationality not to respond to them. No doubt, this, if true, shows us something about the nature of belief, for which an account of belief must offer an explanation. But it will not be the same explanation that helps us to distinguish right and wrong kind reasons in FA theory.<sup>20</sup>

This then concludes the argument for my first claim: the right /wrong kind of reasons distinction in FA theory is different from the distinction which comes under the same name in Reasons Theory.

But this doesn't show that there is a WKR problem, as it is commonly understood in Reason Theory (leaving FA theory behind from now on): a problem that arises for all kinds of reason-sensitive attitudes pointing towards a unified account of theoretical and practical reasons and normativity. I will therefore now pursue the question whether the distinction generalizes: whether there is a similar distinction between reasons for attitudes *other than belief* – in particular whether there is a similar distinction with regard to reasons to intend.

### 3. Reasons to intend and the alleged WKR problem

In the case of belief, I compared truth-related reasons, and reasons that are provided by the value of the attitude of believing. But with intentions, all putative reasons depend on value<sup>21</sup>. Even so, we can distinguish between reasons that are provided by the value of the intended action, and those that are provided by the value of having the attitude (the value of intending). Would this lead to a parallel distinction, a distinction between the right and wrong kinds of reasons to intend?

First of all, it is remarkably difficult to find a parallel to wishful-thinking-type reasons for belief – reasons that are grounded only in the value of having the belief – with

---

<sup>20</sup> A lot more needs to be said about the relation of reasons for belief and rationality here. Assuming that all right kind reasons to believe are evidence-based, is forming a belief rational only if it is based on the available evidence? If not, then the two purported earmarks may apply to different kinds of considerations.

<sup>21</sup> Or at least, more broadly, on normative considerations.

regard to intentions. Imagine that it really matters to me that you should *intend* to come to my party, but it doesn't matter as much whether you actually show up; or an evil demon threatens to torture me unless I *intend* to swallow the contents of a saucer full of mud, but he doesn't care whether or not I swallow it. In both cases, we have reasons to form an intention. We should expect that they are of the wrong kind, since the reason derives from the value of having the intention, rather than the value of acting as intended.

But in these cases, the reasons to intend generate *instrumental* reasons to act as intended: you could simply come to my party, thereby setting my worries to rest, or I could swallow the mud, proving my intention to do so. If we did these things, we would have done them intentionally. The reason for acting as intended would be that there is value in having the intention, and by intentionally acting as we have reason to intend, we have that intention. There is no problem with acquiring the intentions for those reasons, nor would doing so be irrational.<sup>22</sup>

Thus, the fact that there is no independent reason (independent of the value of having the intention) to act as intended doesn't make the reason to intend the wrong kind. By contrast, when there is no independent reason to believe (no reason that is independent of the value of having the belief) the reason to believe would be of the wrong kind. The disanalogy here is not that we can, by acting in a certain way, acquire the intention, whereas we cannot, by acting in a certain way, acquire the belief. We can do the latter too.<sup>23</sup> The disanalogy is that in the case of intentions there is no irrationality in acquiring the intention in this way, whereas in the case of belief there would be.

There is a second important lesson here: the examples above do not seem to be of wrong kind reasons. But the first earmark of WK reasons does nonetheless hold of them: we cannot follow the reason to intend directly. I take it that one follows a reason for an attitude directly if and only if one forms the attitude in response to the reason without taking further steps towards acquiring it. In all those case where we need to take action to acquire an attitude, we do not follow the reason directly. The cases above are of this kind: you follow the reason to intend to come to the party by coming, and I follow the reason to intend to swallow the mud by swallowing it. These are right kind reasons, because they pass according to the second earmark: following the reason (by acting) is rational. But whether or not we count them as RK or WK reasons, the examples show that, in the case of intentions (as opposed to beliefs) the two earmarks pull apart, and don't help to identify a 'kind'.

The only kind of examples that exhibit the two "earmarks" are to be found in the neighbourhood of the toxin puzzle.<sup>24</sup> It seems true that an agent cannot directly

---

<sup>22</sup> For similar observations see Hieronymi (2005), as well as Raz (2009).

<sup>23</sup> We can follow indirect strategies for acquiring beliefs, e.g. expose ourselves only to arguments in their favour, or see a hypnotist.

<sup>24</sup> The toxin puzzle (Kavka 1983): an eccentric billionaire would transfer a million pounds into your bank account at midnight today, if you now intend to drink a (mild) toxin tomorrow. The toxin will not kill you, but it will cause you some discomfort. The billionaire does not require that you drink the toxin, but only that you form the intention to do so. You know now that tomorrow there will be no reason for you to drink the toxin, since either the money is already in your bank account

form an intention to drink the toxin, for the reason that it would get her the money, and if she failed to respond to the reason – that is, if she does not form the intention – the failure does not prove her irrational, even when the reason for having the intention is conclusive and undefeated.

The distinctive feature of this kind of case is that there is (as the agent knows) no reason for acting as intended – not even an instrumental one, as in the examples I mentioned before (i.e. there is no reason to intentionally act as intended as a means to acquiring the intention because in toxin puzzle cases we need to form the intention prior to, and therefore independently of, the intended action).

This may suggest that we cannot rationally form intentions unless we believe that there is some reason for the intended action, instrumental or otherwise. In the party case, you have an instrumental reason for coming to my party, as it proves your intention to do so. So *the wrong kind of reasons to intend* would be those where there is (or: the agent believes there is) no reason to act as intended, but *only* a reason for having the intention. This is rarely the case, because we often can comply with reasons to intend (which are provided by the value of having the intention) by intentionally acting in a certain way. Thus the reason to intend often generates an instrumental reason to act.

If reasons of the toxin puzzle variety are WK reasons to intend, and the ones in the earlier examples are RK reasons, is there any way of saying why which does not draw on the earmarks (which, as we have seen, won't help)? With regard to beliefs we explored the substantive view that evidence for  $p$  is a right kind of reason to believe that  $p$ . In parallel, we may want to say:

**[RKI] A reason,  $R$ , to  $\varphi$  is a right kind reason to intend to  $\varphi$ .**

RKI is a widely accepted claim – tacitly or expressly.<sup>25</sup> The dispute about reasons to intend concerns for the most part the question whether or not there are other reasons of the 'right kind', in addition to those that RKI identifies. So the parallel to the belief-account of the right/wrong kind of reasons distinction could go like this:

1. RKI covers all the right kind reasons to intend. Call this the *exclusivity claim* (borrowing Conor McHugh's apt term<sup>26</sup>).
2. Other putative reasons to intend are of the wrong kind.
3. The explanation of RKI and the exclusivity claim will take roughly the same form as the explanation of the right/wrong kind of reasons distinction regarding belief, thereby showing that practical normativity can be traced back to the nature of intentions in the same way in which epistemic normativity is grounded in the nature of belief.

---

or you won't receive it. Thus you have a reason to form the intention to drink the toxin, but no reason to actually drink it.

<sup>25</sup> Cf. Shah (2008), Schroeder (2012), McHugh (2012), Pink (1991).

<sup>26</sup> McHugh (2013)

However, not only is there no explanation for the truth of RKI that parallels the explanation of why evidence, or truth-related considerations more broadly, are reasons to believe, but RKI is simply false, and consequently we needn't worry about exclusivity – or so I will argue now.

### 3.1. Against RKI

Let me begin by distinguishing between basic and further intentions regarding intentions in action, and then consider whether RKI identifies right kind reasons (or indeed any reasons) for forming intentions of either kind.

Basic intention:  $\phi$ -ing with the intention to  $\phi$ . E.g. I am writing to Paolo with the intention of doing so. Here ' $\phi$ -ing with the intention to  $\phi$ ' contrasts with  $\phi$ -ing mistakenly, or unintentionally, as when I mistakenly send the email which I had intended to send to Paolo to Pat instead.

Further intention:  $\phi$ -ing with the intention of bringing about some result,  $O$ , when  $O$  is different from  $\phi$ -ing. The ordinary use of the phrase ' $\dots$ with the intention to $\dots$ ' is concerned with further intentions, e.g.: I'm writing to Paolo with the intention of inviting him to my party.

Typically, when a person acts with a further intention, she has a basic intention too: when I write to Paolo with the intention of inviting him, I also intend to write to him. And the other way around: when a person acts with a basic intention she often has a further intention: when I write to Paolo, I don't just intend to write to him, but I intend something else: to invite him, or to tell him something. But not all basic intentions are accompanied by further intentions: I may hum a tune intentionally, but with no further intention, not even an intention to get it right.<sup>27 28</sup>

With this distinction in place are right kind reasons to intend reasons to form a basic intention, or a further intention? According to our formulation of RKI, they are presumably reasons for having a basic intention. The reason to  $\phi$  is, RKI tells us, a reason to intend to  $\phi$ , not a reason to  $\phi$  with the intention of bringing about  $O$ , or a reason to have the intention to bring about  $O$  by  $\phi$ -ing.

Now take the example of Julie:

Julie is a talented young author, writing her first novel. There are many reasons for her to do so: that writing the novel furthers Julie's talents, that it will put her on the map as a writer, thereby starting her professional career, that she enjoys writing, or that she is going to tell an important story in a new, and thought-provoking way.

---

<sup>27</sup> Note that in these cases the agent has two intentions: a basic intention and a further one. The claim is not that the basic intention *is* a further one (or the other way around).

<sup>28</sup> There may be further distinctions to be drawn regarding intentions, in particular between intentions in action, and future-directed intentions. Above I explained the distinction between basic and further intentions with regard to intentions in action. But the distinction can be applied *mutatis mutandis* to future-directed intentions as well: I may intend to go for a walk later with the basic intention of going for a walk; or I may intend to go out later to post a letter (further intention).

Might these be reasons to form *further intentions*: i.e. might Julie's reason that by writing a novel she would put herself on the map as a writer be a reason for her to write a novel with the intention of putting herself on the map? They had better not be, for two reasons: first, this way of applying RKI leads to mistakes, as is obvious with the reasons to write because Julie enjoys doing it, or because doing so furthers her talents. These are cases in which having the intention to write in order to enjoy herself, or further her talents, is likely to undermine the result. Secondly, if Julie fails to have an intention to put herself on the map, or to start her professional career, but writes because she is fascinated by the topic she has fastened on, there would be nothing amiss with her intentions. It is not true that Julie's must have a plethora of intentions which reflect the contents of her reasons for acting.

Thus RKI presumably applies in the following way: one of Julie's reason for writing is that she would enjoy doing so (=R). That she enjoys writing is a reason to write, and therefore – applying RKI – a reason to intend to write. And similarly for all the other reasons: they are all reasons for her to intend to write, but not reasons to form any particular further intention in doing so. The content of the intention she has reason to have does not mirror the content of her reasons for writing.

So as we presumed, RKI requires only a basic intention. And intuitively, this seems right: the reason to intend that Julie does have is a reason to intend to write a novel – and this is surely not a reason of the wrong kind.

But getting to this result depends on specifying the reason and the action in the right way: we could think of 'putting oneself on the map' as an action, the reason for doing so being that it is a useful thing to do. In that case, that putting oneself on the map is useful would provide a reason for intending to put oneself on the map. Or the action may be 'furthering one's talents', and the reason for it that doing so is generally a good thing. Then RKI gets us: that furthering one's talents is a good thing is a reason to intend to further one's talents. Thus we'll get a plethora of reasons to intend after all, depending on how we specify Julie's reasons to act.<sup>29</sup> It begins to seem less clear that Julie has all those reasons to intend, even granting that those are reasons for her to act. This is even more obvious if we regard 'enjoying herself' as an activity, and the reason for doing so that enjoying herself is good for Julie. Specified thus, RKI would tell us that Julie's reason to enjoy herself (R = enjoyment is good for her) is a reason for her to intend to enjoy herself. That seems plainly false since forming such an intention is likely to undermine the result. Thus while we can tweak the descriptions of the relevant reasons and actions in such a way that they get us the right result – that Julie has a reason to intend to write a novel – it is not clear why we have to do the tweaking. Why can we not apply RKI to whatever the reason, and the action (or the activity)?

What we are trying engineer is this: Julie has precisely one reason to form an intention, namely a reason to intend to write a novel. She doesn't have a reason to

---

<sup>29</sup> Those different specifications are not in tension, or in competition. Quite commonly, if *R* is a reason to  $\phi$ , there will be some consideration, *Q*, which is a reason to bring about *R*, e.g. if a reason for breaking eggs is that doing so is necessary for baking a cake, there is some other consideration, e.g. that the cake will be tasty, or that I'm hungry, for baking the cake.

do so with any particular further intention, and she doesn't have reasons to intend to engage in any other kinds of action or activity that would lead to complying with her reasons. What explains this?

Here is my hypothesis: a reason to  $\phi$  provides a reason to intend to  $\phi$  only if there is a reason to  $\phi$  intentionally. There presumably is no reason to intentionally further one's talents, or to intentionally put oneself on the map or enjoy oneself. But there is a reason to intentionally write a novel, if one has a reason to write one in the first place. This difference calls for an explanation.

Writing a novel is an activity that determines the pursuit of many other goals by putting them in their place, or banishing them from the productive working hours. It also requires actions other than writing such as doing research on the subject matter, and so on. It is hard to imagine that one could succeed in writing a novel without intending to write one and organising one's life accordingly. Thus having an intention to write is useful (perhaps even indispensable) for succeeding with the activity. This isn't true of all actions but at least of those which are as complex as writing a novel, and can't be done semi-automatically. This kind of reason for having an intention isn't one that could be explained by RKI.

And there is a further reason to reject RKI. If Julie's intention had been to write a short-story which, as it happened, evolved into a novel there would not have been anything amiss with her intentions. She wouldn't have had the wrong intention in this case. Perhaps intending to write a novel would have been daunting for her, and led her to abandon the project whereas the intention to write a short-story provided the right kind of structure and guidance to get into the project of writing a novel which is, let's assume, what she has most reason to do (that's where her talent lies). Thus her reason to write a novel would have given her a reason to intend to write a short story. RKI cannot explain this.

What explains her reasons for having the intention is the instrumental value of having it: having the intention is useful, perhaps even necessary, in order to do something worthwhile like writing a novel. Call this *the facilitative reason to intend*.<sup>30</sup>

It is a textbook case of a reason of the wrong kind (for those who accept RKI), since it is the value of having the intention, the psychological benefit of having it, which provides a reason for it. However, it seems to me that the facilitative reason is the most common reason for forming intentions. Reasons for having a basic intention are, broadly speaking, instrumental: there are such reasons if and because having intentions is sometimes useful to comply with one's reasons.

In addition to having a reason for writing, Julie may have a reason to have a further intention. As we have seen above there are many reasons for acting in a certain way. No doubt, when someone embarks on an enterprise as complex and extended as writing a novel, she is very likely to have a further intention – but it may change over time. Let's assume that at the outset Julie has the further intention of developing the psychology of her main character. The further intention reveals why Julie begins

---

<sup>30</sup> Pink (1991) also argues for the facilitative role of intentions in slightly different way.

writing the novel in the first place: which of the reasons she had for writing it were most salient to her. But her writing for this reason is compatible with complying with the other reasons for writing that she also has, even if she has no particular intention to do so. She would comply with them because (and as long as) they are not reasons to act with a particular intention. She will simply end up developing her talents, starting her professional career, putting herself on the map and enjoying herself. Thus while there is a reason for which a person chooses to undertake a certain pursuit which is reflected in her further intentions, there is no need to mirror all of her reasons to act, or even any particular one, in this way.

There are, however, reasons other than the facilitative reason to do something intentionally: some actions have value only if they are done with (or without) a particular intention. Take 'thanking someone' as an example. When there is a reason for thanking a person, this is *ipso facto* a reason for thanking her intentionally. An action that is not done with the intention of thanking another may not even count as thanking her. But whether or not it does, it wouldn't have value as an expression of gratitude. There is value in this action only when it is done with the right intention. A reason for an action of this kind would therefore always be a reason to have (or in different cases: lack) the relevant basic intention. But it is not in general true that the value of an action depends on its being done with a particular intention.

Furthermore, even in these cases, the reason to intend does not derive from the reason to act in the way RKI has it. The reason to act is, from the start, a reason-to-act-with-intention-I. There is no further (derivative) reason to *intend* to act with intention I. Thus, in those cases the reason to intend is not explained by RKI.

What explains our reasons to intend is not RKI. We have a reason to form an intention when we have a reason to do something intentionally. Sometimes, complying with a reasons requires acting intentionally, because the intention (some intention, at any rate) is needed for successfully completing the action (the facilitative reason), or because the value of the action depends on the intention with which it is done.

This then is my case for the second claim: the claim that there is no WKR problem which takes the same form for reason to intend and reasons to believe.<sup>31</sup> There is no unifying explanation of a shared phenomenon (the W/RK distinction) that could lead to a unified account of theoretical and practical normativity, since there is no such phenomenon. Not only is it false that reasons to  $\phi$  are RK reasons to intend to  $\phi$  – they are, all by themselves, no reasons for forming any intention. The most common reason to intend – the facilitative reason – is explained by the value of intending, and therefore seemingly of the wrong kind.

### 3.2. Toxin

There remains one loose end which I would like to tie up: the Toxin puzzle illustrates that we cannot, or cannot rationally, intend to  $\phi$ , unless there is (as we see it) a reason to  $\phi$ . It is perhaps this observation that lends RKI whatever initial plausibility

---

<sup>31</sup> For a more comprehensive discussion see my 'Reasons to Intend' (forthcoming).

it seems to have. So if we reject RKI, why then is it that we can't follow the reason to intend in the Toxin Puzzle? After all, having the intention would be very useful in this case too. Isn't there a facilitative reason for forming it then?

If a person were to form an intention to drink the toxin, she would (if she is rational and reflective) do so, knowing that she will not act on it. There is a long-standing discussion<sup>32</sup> whether having an intention involves a belief, or even more strongly: certainty, that one will act as intended. If it did, solving the toxin puzzle would be easy: a rational agent would know that she is not going to act as intended, since she would know that, at the time of action, there will be no reason for her to do so. Hence she cannot form the intention. However, the assumption that having an intention involves a belief, or a certainty, of this kind is doubtful. We are able to form intentions when we are not sure whether or not we will act as intended. New Year's resolutions may be a case in point: there would be no need to even form such a resolution if we already believed, or were certain, that we will act as we resolve to do. But a sincere resolution of this kind involves an intention to act on it.<sup>33</sup>

Yet even if forming an intention does not require the belief that one will act as intended, it may require the absence of the belief that one will not do so. This, I think, explains the Toxin puzzle: you cannot form the intention to drink the toxin, because you know ahead of time that you will not drink it (again: assuming that you are reflective and rational). Thus an agent can – reflectively and rationally – form an intention to  $\phi$  only if she does not believe that she will not  $\phi$ . This, if true, calls for a further explanation, and it certainly does show us something about the nature of intentions. I will not pursue the issue here, but it seems clear at the outset that the explanation will have little to do with the explanation of the distinction of WK/RK reasons to believe, since there is no commonality in the *explanandum*.

The facilitative reason to form an intention is, by contrast, a reason to form an intention because doing so helps to comply with one's reasons to act. Thus, when there is a facilitative reason there is always, necessarily, a reason to act as intended as well. This contrast explains why the facilitative reason is a standard case of a reason to intend, rather than a WK reasons, despite the fact that it is provided by the value of having the intention.

### Conclusions

First, the right/wrong kind of reasons distinction in FA theory is independent and different from its alleged uses elsewhere.

Secondly, beyond FA theory, there is a distinction with regard to reasons for belief: between evidence for the truth of  $p$  (or more broadly: truth-related reasons for

---

<sup>32</sup> E.g. Grice (1972), Harman (1976), and, more recently, Setiya (2012) all argue for different versions of this view.

<sup>33</sup> Resolving is different from intending, and perhaps having resolved to  $\phi$  strengthens the likelihood that you will  $\phi$ . If so there may be a stronger reason to believe that you will  $\phi$  because you have resolved to do so. But even so, there is nothing like certainty here.



believing  $p$ ), and reasons to believe  $p$  because having the belief would be good in some respect, independently of whether or not  $p$  is true. Reasons of the latter kind exhibit two earmarks: we cannot follow them directly (we cannot reason from them to believing  $p$ ), and failing to comply with them is not a failure of rationality. While there is a striking contrast between these kinds of reasons to believe, there is no similar distinction regarding reasons to intend. There are two ways of drawing the distinction: (1) by appealing to the two earmarks; or (2) by giving a substantive account: only evidence-based reasons are right kind reasons for belief. They converge on the same distinction. With regard to reasons to intend, neither one works. Regarding the first, some right kind reasons bear the first earmark of wrong kind reasons: that we cannot follow them directly. The two earmarks pull apart. With regard to the substantive account, I considered the generally accepted view that reasons to act provide right kind reasons to intend (similar to the role of evidence in the epistemic case). It turned out that in many cases reasons to act don't provide any reasons to intend, and that the most common reason to intend, the facilitative reason, is provided by the benefit of having an intention, not by the reason to act. The hypothesis I defended is that, if there is a sufficient reason for doing something intentionally, then there is a reason to form an intention.

Thus there is nothing on the RK side of the alleged distinction that resembles RK reasons to believe, and there is nothing on the WK side either. What makes Toxin puzzle reasons 'wrong' (this was the only remote relative of WK reason I could make out on the intention side, going by the two earmarks as their distinguishing feature) is very different from whatever explains the distinction with regard to reasons to believe.

The remaining truth is that in order to understand reasons for having particular kinds of attitudes, we have to understand the nature of those attitudes. But there is no distinction between the right/ wrong kind of reasons that bears a rough similarity across attitudes, and could contribute to understanding the rationality of those attitudes along similar lines. It therefore seems to me that we would do well to abandon the talk of the right and wrong kind of reasons outside of FA theory altogether!<sup>34</sup>

### References

- Grice, H.P. 1972: 'Intentions and Uncertainty'. *Proceedings of the British Academy*, pp. 263–279.
- Harman, Gilbert 1976: 'Practical Reasoning'. *Review of Metaphysics*, 29 (3), pp. 431–63.

---

<sup>34</sup> I presented early versions of this paper at a workshop in honor of Wlodek Rabinowicz in Lund, at a workshop on the Wrong Kinds of Reasons at the Humboldt University in Berlin and to the Interuniversity Normativity Research Group at Montréal, and at a workshop on Epistemic and Practical Normativity at Southampton. I'm grateful to these audiences for helpful discussions and comments – in particular to my commentator at Southampton, Chris Howard. Two anonymous referees prevented a number of blunders. Finally, I owe a special debt of gratitude to Wlodek Rabinowicz for his sharp, constructive and encouraging response to a very early version of this paper.

- Heuer, Ulrike 2006: 'Explaining Reasons: Where Does the Buck Stop?' in *Journal for Ethics and Social Philosophy* 1(3): 1-25.
- 2011: 'Beyond Wrong Reasons: The Buck-Passing Account of Value'. Michael Brady (ed.), *New Waves in Metaethics*, Palgrave, pp. 166-184.
- (forthcoming): 'Reasons to Intend' in Daniel Star (ed), *The Oxford Handbook of Reasons and Normativity*. Oxford University Press.
- Hieronymi, Pamela 2005: 'The Wrong Kind of Reason'. *Journal of Philosophy*, 102 (9), pp. 437–57.
- Kavka, Gregory 1983: 'The Toxin Puzzle'. *Analysis*, 43, pp. 33-6.
- Lang, Gerald 2008: 'The Right Kind of Solution to the Wrong Kind of Reasons Problem.' *Utilitas* 20 (4), 472-89.
- McHugh, Conor 2011: 'What Do We Aim at When We Believe?'. *Dialectica* 65(3): 369-392.
- 2012: 'Control of Belief and Intention'. *Thought*, 1 (4), pp. 337–46.
- 2013: 'The Illusion of Exclusivity'. *European Journal of Philosophy*, 22 (3). doi: 10.1111 / ejop.12032.
- 2014: 'Fitting Belief'. *Proceedings of the Aristotelian Society*, 114 (2), 167-187.
- McHugh, Conor and Jonathan Way (forthcoming): 'Fittingness First'. *Ethics*.
- Olson, Jonas 2004: 'Buck-Passing and the Wrong Kind of Reasons'. *Philosophical Quarterly* 54, 295-300.
- Owens David 2003: 'Does Belief Have an Aim'. *Philosophical Studies* Vol. 115, No. 3, pp. 275-97.
- Pink, T. L. M. 1991: 'Purposive Intending'. *Mind*, 100, pp. 343–59.
- Rabinowicz, Wlodek & Toni Rønnow-Rasmussen (2004): 'The Strike of the Demon: On Fitting Pro-Attitudes'. *Ethics* 114, 391-423.
- 2006: 'Buck-passing and the Right Kind of Reasons'. *Philosophical Quarterly* 56, 114-120.
- Raz, Joseph 2009: 'Reasons: practical and adaptive'. In D. Sobel and S. Wall (eds.), *Reasons for Action*, Cambridge: Cambridge University Press, pp. 37-57.
- Scanlon, T.M. 1998: *What We Owe to Each Other*. Cambridge (Mass): Harvard University Press.
- Schroeder, Mark 2012: 'The Ubiquity of State-Given Reasons'. *Ethics*, 122 (3), pp. 457–88.
- 2010: 'Value and the Right Kind of Reasons'. *Oxford Studies in Metaethics* 5, 25-55.
- Setiya, Kieran 2004: 'Explaining Action'. *Philosophical Review*, 112, pp. 339–93.
- 2007: *Reasons Without Rationalism*. Oxford: Oxford University Press.
- 2012: 'Knowing How'. *Proceedings of the Aristotelian Society*, Vol. CXII, part 3, pp. 285-307.
- Shah, Nishi 2008: 'How Action Governs Intention'. *Philosophers' Imprint*, 8 (5).
- Way, Jonathan 2012: 'Transmission and the Wrong Kind of Reason'. *Ethics*, 122, (3), 489-515.

