PRIMARY AND SECONDARY REASONS

Ingmar Persson

Department of Philosophy, Gothenburg University

Ingmar.Persson@phil.gu.se

ABSTRACT:This paper argues that Wlodek Rabinowicz and Toni Rønnow-Rasmussen are wrong in thinking that what they call the 'Wrong Kind of Reasons' problem presents a serious problem for the idea that the fact that there are reasons to have a pro-attitude towards an object implies that it is valuable. It seems a serious problem to them because they mistakenly reject the view that some reasons that in everyday language are described as reasons for an attitude are really reasons for wanting, intending or trying to have it. This view is here defended against Rabinowicz' and Rønnow-Rasmussen's attack by an account of what attitude a reason is a reason for in terms of what direct response it justifies as an outcome of a piece of reasoning.

The Wrong Kind of Reasons Problem

In their paper 'The Strike of the Demon: On Fitting Pro-attitudes and Value' Wlodek Rabinowicz and Toni Rønnow-Rasmussen press a certain objection against attempts to define the value of something in terms of properties of this thing providing reasons — or making it fitting — to have a pro-attitude, such as a desire, towards this thing — the FA analysis (short for 'fitting attitude' analysis).[1] They call this objection the 'Wrong Kind of Reasons' problem, the WKR problem (p. 393). I shall in this paper examine, and dismiss, their rejection of a solution to this problem. Thus, I believe that the WKR problem can be solved.

I think, however, that the FA analysis to which it is presented as a problem faces other difficulties. Rabinowicz and Rønnow-Rasmussen question the implication from there being reasons to adopt a pro-attitude towards a thing to it being valuable. I believe the implication in the other direction — to the effect that if something is valuable, there is something about it that provides a reason to adopt a pro-attitude towards it — is more problematic. For it seems to me doubtful, e.g., that there is something about pleasure that provides a reason to want it (that pleasure is pleasure appears to be no such reason), though pleasure is (intrinsically) valuable. But in this paper I shall not attempt to settle whether or not the FA analysis is sound; my aim is just to show that Rabinowicz and Rønnow-Rasmussen are mistaken when they turn down a certain answer to the WKR problem.

[1] *Ethics*, 114 (2004), pp. 391-423. Unprefixed page references are to this paper.

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

Rabinowicz and Rønnow-Rasmussen agree that when it is valuable to have a pro-attitude, 'we do have reasons to want it or to try to have it', but they claim that then 'we have also reasons to have' it (p. 412). I call this *the downward claim*:

(DC) If anything, R, is a reason to want (to have) a pro-attitude, e.g., a desire for X, (and to try to have this attitude), R is also a reason to desire X (and to try to get X).

In other words, a reason to have a higher-order pro-attitude which has a lower-order pro-attitude as its object or content is also a reason to have the lower-order attitude. Thus, the force of reasons extends 'downwards', from higher-order pro-attitudes to lower-order ones (or at least from second-order to first-order attitudes). (DC) is a nuisance to the FA analysis. For if it is true, the fact that there are reasons for having a higher-order attitude implies that there are reasons for the lower-order attitude which forms its object and therefore, according to the FA analysis, that the object of the lower-order attitude is valuable. But that implication seems patently false.

Rabinowicz and Rønnow-Rasmussen borrow an example of Roger Crisp's to illustrate the point (p. 402). Suppose an evil demon now threatens to inflict severe pain on you if you do not desire a saucer of mud. This threat (if credible) provides you with a reason to want to have a desire for a saucer of mud. Thus, according to the FA analysis it is valuable to have this lower-order desire. But if (DC) is true, it follows that you have also a reason to desire a saucer of mud and, so, that the saucer of mud is valuable – which is obviously preposterous.

It seems quite natural to say that you have a reason to desire a saucer of mud in this case, but appearances are deceptive. Imagine that the demon's threat had been a trifle more precise: to inflict severe pain on you if you do not desire a saucer of mud *on Friday next week*. This gives you a reason now to want (now) to have a desire for a saucer of mud on Friday next week. According to (DC), it also gives you a reason now to desire a saucer of mud on Friday next week, but it is not clear what this could mean if it means anything different from the foregoing. It cannot mean that you now have a reason to want *now* to have a saucer of mud on Friday next week, for the demon's threat provides you with no such reason: it supposed to provide you with a reason now to have this desire on Friday next week, not now. But it seems that you cannot now have a reason for having an attitude at a future time so distant as Friday next week, for this sort of temporal disconnection could prevent you from forming the attitude on the basis of your reason. (Suppose the demon will give you an amnesiac sometime before Friday next week; then you cannot on that day form your desire on the basis of a reason you now have.) Consequently, it is not clear to me what intelligible meaning (DC) could have in cases in which you allegedly have a reason for having an attitude at a time more distant than the immediate future. But henceforth I shall put such cases aside and focus exclusively on cases in which you are supposed to have a reason for having an attitude 'without delay'.

Contrast (DC) with *the upward claim*:

(UC) If anything, R, is a reason to have a pro-attitude, e.g., a desire for X, R is also a reason to want to have the desire for X.

That is to say, the force of reasons extends 'upwards', from lower-order to higher-order attitudes. It may be less clear that Rabinowicz and Rønnow-Rasmussen endorse (UC) than (DC). But they seem to do so, for they claim that we have reasons for higher-order attitudes 'because' we have reasons for the lower-order attitudes which form their objects (p. 413). Furthermore, if they do not endorse (UC) as well, they invite a reply to the WKR problem. For if there are reasons to have a pro-attitude which are not reasons to want to have it, it could be claimed that, according to the FA analysis, it is these and only these reasons for a pro-attitude which make the object of the pro-attitude valuable. It would then be irrelevant if reasons for wanting to have the pro-attitude are also reasons for the pro-attitude and that it would be absurd to claim that these reasons make the object of the lower-order attitude valuable. For the FA analysis would not cover these reasons for a pro-attitude which are transferred from a higher-order attitude.

On the other hand, there are also reasons not to embrace (UC) alongside (DC). (DC) and (UC) together make up *the combined claim*:

(CC) Something, R, is a reason to have a pro-attitude, e.g., a desire for X, if and only if

R is also a reason to want to have the desire for X.

But (CC) makes one wonder how desiring X and wanting to desire X can be different pro-attitudes, for it might seem that we would have two ways of identifying one and the same attitude if whatever is a reason for one attitude is a reason for the other and *vice versa*. Moreover, (UC) by itself is problematic. It seems clear that there are reasons for wanting to do some things which are not reasons for wanting to have a desire to do them. Thus, if a strong desire to do risky things (in order to show off my courage) is among my character traits, this gives me a reason for wanting to do (and doing) certain acts merely because they are risky. But the fact that some acts are risky gives me no reason for wanting to have a desire to do them (as a display of courage). It seems more of a reason *against* wanting to have a desire to do them.

Furthermore, it would be rather odd if a reason for desiring X was also a reason wanting to have this desire. For if this reason were sufficient to generate a desire to have a desire for X, it would presumably be sufficient to generate a desire for X. So to have the higher-order desire would be to desire to have what is already had, and this seems impossible (assuming that there is knowledge that the desire is had). For this is a situation in which satisfaction rather than desire is appropriate. Thus, the reason for the higher-order attitude would be one for which we would never have any use. If it is replied that there could be reasons which are practically useless, it should be noticed that (UC) invites not a few, but an infinite hierarchy of such useless reasons. For (UC) raises the question why reasons for wanting to have the desire for X are not also reasons for wanting to have the desire to have the desire for X, and so on. This seems to be ontological extravagance.

At first sight, it might be thought that (DC) is vulnerable to similar objections. For if it is true that, whenever you have a reason to form a higher-order desire to have a desire, you have a reason to form the lower-order desire that is its object, it might seem to follow that the higher-order desire would be

automatically fulfilled by your having the lower-order desire. But this need not be so if (UC) is false, and there could be reasons for the lower-order desire which are not reasons for the higher-order desire. For this kind of reasons could block the birth of the lower-order desire and, thus, leave the higher-order desire, which they would not block, unfulfilled. On the other hand, as I have remarked, the denial of (UC) opens the door to a reply from FA analysts to the WKR problem. Thus, Rabinowicz and Rønnow-Rasmussen face something of a dilemma.

Primary and Secondary Reasons

Having sketched this broader dialectic picture, I shall leave the question of the truth of (UC) aside and concentrate upon (DC), as Rabinowicz and Rønnow-Rasmussen themselves do. In favour of (DC), it must be admitted that, in everyday parlance, we do talk as though it were true. For instance, it would be quite natural to say, about Crisp's example, that the demon's threat is a reason to desire a saucer of mud. But, on reflection, it is undeniable that wanting to have a desire for X is a different desire than desiring X just as believing that one believes *p* is a different belief than believing *p* and, since it is clearly not the case that what is a reason for believing that one believes *p* is perforce a reason for believing *p*, the question arises why the situation should be different as regards desires, so that reasons for wanting to have a desire for X would perforce be reasons for desiring X.

To find out whether (DC) true, it is useful to first consider a case in which there is reason to want to have a belief *p* and ask whether it is also a reason to believe *p*. Consider Pascal's wager who is given the reason to want to believe that God exists that, if he has this belief, he will be amply rewarded after death (which is a proper object of desire).[2] Now, we can certainly *say* that this is a reason to believe that God exists. But surely this manner of speaking is misleading (and so the parallel manner of speaking manifested in (DC) might be misleading, too). For this reason for believing that God exists is clearly different from a reason for this belief that supports its truth or makes it more probable than it earlier was. Let us call the latter kind of reasons *epistemic* reasons for belief and the former sort *pragmatic* reasons for belief.

In itself, this distinction does not give us any inroads on (DC) for there are no epistemic reasons in the case of pro-attitudes, since they are not designed to be true. It may however help if we dig deeper into the distinction between epistemic and pragmatic reasons for belief. For it may be that we then discover something which has a counterpart in a distinction between two kinds of reasons for pro-attitudes.

As Pascal himself noticed, we cannot *directly* respond to his pragmatic reason for believing in God's existence by believing this: we can just act as if we believed this, attend mass and so on. In

[2] *Pensées*, Oxford: Penguin, 1966 (1670), pp. 149-53.

contrast, we *can* directly respond to an epistemic reason for this belief — e.g., a reason to the effect that nature shows clear signs of a pervasive intelligent design — by forming the belief. We can directly respond to epistemic reasons for believing *p* by believing *p* because we see such reasons as *justifying* the belief *p* by supporting its truth. The fact that we can directly respond with a belief to something we take to justify the belief requires no explanation, as long as we are allowed to take it for granted that we can function rationally. Thus, the epistemic reasons we have for believing *p* have the definitive function of directly producing — or supporting — our belief *p* by being seen by us as justifying this belief (as true).

We have epistemic reasons that are seen by us as (directly) justifying our belief *p* if and only if our belief is the (immediate) conclusion of a piece of reasoning that we believe to be valid (though it need not be valid in fact) and whose premises we believe to be true (though they need not be true). For instance, the belief that nature shows clear signs of a pervasive intelligent design is a possible epistemic reason for believing that God exists, since in conjunction with the possible belief that if nature shows clear signs of a pervasive intelligent design, God exists, it makes up a credible piece of reasoning whose conclusion is that God exists.

In contrast, you cannot directly respond to the belief that if and only if you believe that God exists, you will be rewarded with eternal bliss, by believing that God exists because you see the second belief as justified by the first belief. For you are not likely to see the first belief as a part of a reasoning whose (immediate) conclusion is the second belief. Your direct attitudinal response — i.e., the attitude you form without first forming any other attitude — to the first belief is more likely to be a desire to believe that God exists. For if you desire eternal bliss, as you may reasonably do, you will see the first belief as justifying a desire to believe in God's existence. That is to say, there is piece of — practical — reasoning which you endorse and which directly leads to this conclusion.

So, although the belief that if and only if you believe that God exists, you will be rewarded with eternal bliss can be said, in everyday talk, to be a reason for believing in God's existence, it should be distinguished from epistemic reasons for the latter belief. I have proposed to do so, terminologically, by calling it a pragmatic reason for belief. I now claim that a pragmatic reason for a belief is strictly speaking a reason for wanting to have this belief. For you can directly respond to the belief composing such a reason — e.g., the belief that if and only if you believe that God exists, you will be rewarded with eternal bliss — by having this desire because you can see this belief as justifying this desire, i.e., as forming (with a desire for eternal bliss) a practical reasoning of yours which directly issues in the desire to believe in God.

If you have this desire to believe in God, you may try to fulfil it by looking around for epistemic reasons in favour of the belief that God exists, by suppressing epistemic reasons against it, and so on. As a result, you may acquire this belief. Pragmatic reasons for belief may therefore characterized as *secondary* reasons for belief as opposed to epistemic reasons which are *primary* reasons for belief. They are secondary because they generally need to be assisted by or supplemented with epistemic

reasons in order to produce the relevant belief (but the reverse does not hold). This is because a belief cannot be directly produced by a desire to have it as a bodily movement often can be.

There is a reason why beliefs are not under this sort of voluntary control: it would not sit well with the function of beliefs to fit the facts. Beliefs cannot have this function because we want to have beliefs that fit the facts, for if some of our beliefs did not have another source than this desire of ours, we could not even begin to satisfy this desire. Beliefs can however have this function if their standard source is that they are caused by the facts, through some generally reliable mechanisms, or by other beliefs having this source. But given the existence of such mechanisms, we could on occasion employ them to instill in ourselves beliefs that we want to have (e.g., by changing the facts so that the beliefs become true). In this way, beliefs can indirectly be under our voluntary control.[3]

The reason why we do not have direct voluntary control over our beliefs is probably that it would interfere with this factual grounding. But it is noteworthy that even if our beliefs were under our direct voluntary control, they would still not be direct responses to pragmatic reasons, for they would nevertheless be produced *via* the intervention of *another propositional attitude*, a desire to have them. This supplies another reason for describing pragmatic reasons for beliefs as secondary reasons, for they are really reasons for a higher-order (conative) attitude which takes belief as its object.[4]

But could not a belief that *p* conceivably be a direct *non-voluntary* response to a belief that it would be good to have the belief that *p*? Certainly, we cannot *a priori* exclude all sorts of wayward causation. But in order for the former belief to be a primary reason for the latter belief, the latter must be construable as a *justified* direct response to the belief that it would be good to have this belief. However, we could not construe beliefs as being justified by it being good to have them as long as beliefs remain states designed to fit the facts.

---

[3] In an unpublished manuscript, 'Responding to Reasons', Derek Parfit makes similar claims.

[4] The distinction I have drawn between primary and secondary reasons seems be the same as the distinction Pamela Hieronymi draws between 'constitutive' and 'extrinsic' reasons ('The Wrong Kind of Reason', *Journal of Philosophy*, CII (2005), pp. 437-57). Constitutive reasons for believing *p* bear on the question whether or not *p* is true. She calls such reasons 'constitutive' because they '(are taken to) support the commitment that is constitutive of the attitude' (p. 449). In the case of the belief *p* this is commitment to the truth of *p*. It is because these reasons bear on whether or not *p* is true that we can directly respond to them by believing *p*, and to say that they are taken to support the truth of *p* is another way of saying that we see this response as justified. I am however uneasy about the notion of 'the commitment that is constitutive of the attitude' in the case of some pro-attitudes. If there is such a commitment in the case of desire, it appears to be a commitment to *the desirability* of the object. But if there is a reference to desirability, it may seem circular to rely upon the concept of constitutive reasons for desire in replying to the WKR problem as regards FA analyses.

It should be admitted that the notions of justification and of a valid pattern of reasoning are better understood and, so, more uncontroversial in the theoretical case, in which the conclusion is a belief, than in the practical case, in which the conclusion is a pro-attitude like a desire. An epistemic reason for a belief justifies this belief to the extent that it supports the truth of this belief or renders it probable. A valid piece of theoretical reasoning is one that is truth-preserving: its conclusion is true if its premises are. Thus, in the theoretical case justification and validity can be elucidated by appeal to the concept of truth. Of course, this interpretation will not do in the practical case, since the conclusion is here something, e.g., a desire, which is not designed to fit the facts, like a belief. For this reason, it is inappropriate to speak of epistemic reasons here. Instead, I shall simply speak of primary and secondary practical reasons.

Although the notions of practical justification and valid practical reasoning are more controversial, we still have some intuitive understanding of these notions. Thus, we recognize a reasoning from the belief that one will gain eternal bliss if and only if one believes that God exists and a desire to gain eternal bliss to a desire to believe in God as a candidate of valid practical reasoning. If the belief is true and the first desire is justifiable[5] then the second desire is justifiable, too. This is enough to give substance to the claim that we have here a primary reason for desiring to believe in God's existence just because we can directly respond to it by this desire because we see the reason as justifying this desire.[6]

With this in place, we can return to Crisp's case of the evil demon. It is now apparent that the belief that you will be severely punished by the demon if you do not desire a saucer of mud is a secondary reason to have this desire, i.e., it is a primary reason to have the higher-order attitude of wanting to have this desire. For assuming that you reasonably want to avoid severe punishment, you can directly respond to this belief by wanting to have a desire for a saucer of mud because you see it as justified. But you could not directly respond to this belief by desiring a saucer of mud because you see this desire as justified. For, in all probability, there is no practical reasoning attributable to you which has

[5] There are different accounts of the ultimate justification of desires, e.g., externalist or internalist ones. But there is no need to take a stance in this debate for the purposes of this paper. It is desirable that a reply to the WKR problem be neutral between these accounts.

[6] In 'Responding to Reasons', Parfit uses the expression 'directly respond' to describe the formation of belief and desire on the basis of reasons. But it should be stressed that the requirement of the response being direct is not by itself sufficient. For I may directly respond to your statement that you just had an ice-cream by desiring to have one myself. This does not make your statement my reason for having an ice-cream. It merely puts the idea of ice-cream in my head and, so, makes me think of the delights of having one. These delights are what I see as justifying my desire; thus, they provide the content of my primary reason for the desire.

this desire as its conclusion and the belief as its premise. Therefore, the belief that you will be punished for not desiring the saucer of mud is not a primary reason for desiring it.

Again, as in the case of belief, if you have (primary) reasons for wanting to have a desire, such as the desire for a saucer of mud, you will have to cast about for primary reasons for the latter desire in order to be able to fulfil the former desire. For our desires are as little under our voluntary control as our beliefs are: we do not directly control our desires as we control some bodily parts. It is logically impossible that we always have our desires because we want to have them: we would then be caught up in an infinite regress. So, at least some of them must have a non-conative source. There is an evolutionary explanation of why this source should be our beliefs which often enough are caused by the facts. Once this mechanism is in place, we can employ it to try to cause ourselves to have desires that we want to have. Thus, in this case as in the case of belief, secondary reasons for a desire need to be supplemented with primary reasons for the lower-order desire in order to produce it.

The reason why we do not have a capacity to produce directly desires is presumably that it would interfere with the operation of the doxastic ground. But it should still be noticed that, even if you could directly produce, say, a desire for the saucer of mud by an act of will, without having any primary reasons for it, it would still not be a direct response to the reasoning outlined in the foregoing paragraph, since it would be mediated by a desire to have this desire which *would* be a direct response to the reasoning.

Suppose, contrary to this reasoning, that (DC) were true. Then if your reason to want to have the desire for a saucer of mud were strong enough to produce this higher-order desire — which it could easily be as the punishment is severe — it seems it should be strong enough to produce the desire for a saucer of mud. Thus, you would not need to cast about for any other reasons in order to have this desire. But surely it is not as easy as this to acquire the desired lower-order desire.

In fact, this seems no easier than it is for Pascal's wager to acquire a belief in God's existence for the reason that it would be good for him. If reasons for wanting to believe were transformable into reasons for believing, Pascal's wager should be able to acquire easily the required belief, since he would have a very strong reason for it because he has a very strong reason for wanting to have it. But wishful thinking is not as easy as that. This is because reasons for wanting to believe strictly speaking are not reasons for believing, i.e., they are not epistemic reasons. Analogously, wishful wanting is not easy because reasons for wanting to have a desire are strictly speaking not reasons for having the desire. To fulfil your higher-order desire, you need to cast about for other (primary) reasons for the lower-order desire.

By now you have presumably guessed what my answer to the WKR problem is: the wrong kind of reasons are secondary reasons; the value of something should be defined by FA analysts in terms of there being primary reasons for a pro-attitude towards this thing. Thus, we avoid the absurd corollary that a saucer of mud is valuable, since your reason for desiring it is secondary, i.e., it is a (primary)

reason for wanting to have this desire. Therefore, it is only your having the desire for the saucer of mud, not the saucer of mud, which is (instrumentally) valuable, according to the FA analysis.

At the end of their critique of the sort of approach I have advocated, Rabinowicz and Rønnow-Rasmussen suppose, for the sake of the argument, that this approach is right. They then remark that those who advance it would

> 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 towards an object rather than a direct reason for having the attitude in question (pp. 413-4).

I have tried to clarify this in terms of what attitude could be a direct response to the reason because it is seen as justified by it. This clarification holds good irrespective of whether the justified attitude is a pro-attitude, like a desire or emotion, or a belief, though there is admittedly a better understanding of the notion of justification (and valid reasoning) in the case of belief.

Objections

In a superficial sense I have conceded the truth of (DC), for I have conceded that in loose, everyday language, a reason for wanting to have a pro-attitude can be said to be a reason for having the pro-attitude. But in a stricter sense, I have rejected (DC) for I have denied that reasons for wanting to have a pro-attitude can *in the same* (primary) *sense* be reasons for the pro-attitude. Surely, the latter must be the sense in which Rabinowicz and Rønnow-Rasmussen take (DC).

They seem to think that the rejection of this claim issues in the counter-intuitive result that one could find oneself vainly desiring not to feel emotions that one has good (primary) reasons for having, e.g., a courtier, who is afraid to enrage the queen, vainly desiring not to feel angry at the queen when he has good reasons for this emotion and so is driven to feel it (p. 412). If, in accordance with (DC), the strong reasons for desiring not to feel the anger had instead translated into equally strong reasons against feeling the anger, the anger could have been prevented by the fact that the reasons for it are outweighed by stronger contrary reasons. However, the implications of my view for this type of case are not these counter-intuitive ones, but are rather borne out by everyday experience. If the reasons for desiring not to feel the anger are strong enough, and the desire consequently is strong, it is likely not to be in vain. For it would then animate the courtier to make determined efforts not to think of the (primary) reasons for the anger and, so, he would probably succeed in not thinking of them and, thus, not feel the anger. This is a well-known recipe for avoiding to feel emotions that one does not want to feel: try thinking of something else than the reasons for the emotions. I have already touched upon this kind of recipe in connection with Pascal's wager and the fostering of beliefs.

Rabinowicz and Rønnow-Rasmussen rhetorically ask why the fact that it is valuable to have an attitude cannot be a reason for having the attitude when the fact that it is valuable to do an action can be a reason for doing the action (p. 413). The answer is that a reason for doing an action is a reason for *wanting*, *intending* or *trying* to do it in circumstances in which you can do it. Thus, every reason for doing an action is a reason for wanting, etc. to do it, but you can have a reason to want, etc. do it without it being true that you have any reason to do it, because you cannot do it ('ought' implies 'can'). Now, if the value of an action is a reason for wanting, etc. to do it, this is parallel to the situation in which the value of an attitude is a reason for wanting, etc. to have it. We do actions for reasons only in so far as they are manifestations of desires or intentions, i.e., only in so far as we (successfully or unsuccessfully) *try* to do something in doing them. We do not do actions for reasons when they are done on reflex, like a movement to catch something that suddenly falls, or absent-mindedly, like tapping one's feet to the rhythm of music. For although these actions are caused by beliefs we have, we do not act in these ways because we see our actions as justified by our beliefs.

Compare with the situation with respect to our reasons for or against performing basic actions and the non-basic actions which we perform by means of the basic actions. To recycle an example Donald Davidson used in an early, classic discussion: I turn on the light by flipping the switch and I am alerting the prowler by turning on the light. The undesirability of alerting the prowler provides me with a reason both against flipping the switch and turning on the light. Do I then have two reasons? No, for there is only one reason-providing fact — the undesirability of alerting the prowler — and it is a reason against flipping the switch because the light will thereby be turned on, and the prowler alerted. This explains why I have failed to respond to the reasons I have, and have been equally irrational, irrespective of whether the light is turned on when I flip the switch. Analogously, my reason against flipping the switch is a reason to try to avoid flipping it. It can be redescribed as a reason against trying to flip the switch because I shall then succeed in not flipping it and not turn on the light and alert the prowler. If I try to avoid flipping it, but fail because your stronger hand forces me to flip it, I have not failed to respond to a reason I have and exhibited some measure of irrationality.

On the other hand, I can have reasons against trying to flip the switch which are not reasons against flipping the switch, just as I can have reasons against flipping the switch which are not reasons against turning on the light, for I can do the former in circumstances in which I cannot do the latter, and the former can have undesirable consequences which do not occur via the latter. For instance, if I am threatened with severe punishment if I try to square the circle which, unbeknownst to me, is impossible, I have strong reason not to try this, but I cannot have any reason not to square the circle, since this cannot be done ('ought' implies 'can'). The unpleasant consequences which provide me with a reason occur as soon as I try and do not depend upon the success of my attempt.

This excursion into the theory of action enables me to reply to another objection to the sort of account I have given of primary reasons. Thomas Kelly argues that the directness of the response cannot be essential because one could have the requisite sort of reason for a non-basic action, though

one cannot directly respond to this reason by performing the non-basic action: by definition, this action has to be performed indirectly, through some basic action.[7] But, as remarked, the undesirability of alerting the prowler is my reason not only for not turning on the light, but also for not flipping the switch — and, indeed, trying to avoid this — and this is in my repertoire of possible direct responses.

It might however be thought that there are counter-examples to the claim that having a reason for doing an action is having a reason for wanting, intending or trying to do it because wanting, intending and trying to do an action can be unnecessary for doing an action that you seem to have reason to do. Suppose, for instance, that you come to believe that you have a reason to see to it that you get food because you realize that you are undernourished, but then you immediately notice that you are being force fed. Then you seem to have no reason to want, intend and try to see to it that you get food because you are already getting it. But if the situation is this, what you have a reason for changes: the reason you have is transformed from a reason to see to it that you get food to a reason to see to it that you *continue* to get food or that you not resist the force feeding or some such. Now, these are also things that you have reason to want, intend and try and, furthermore, they are things you can do for your reason. Thus, I believe, such discrepancies between reasons for action and reasons for wanting, intending and trying are only apparent.

Generally speaking, what we have primary reasons for or against is adopting, retaining or abandoning *propositional attitudes*, since it is only states with propositional content which can conclude (as well as supply the premises of) a piece of reasoning.[8] Desires and emotions have propositional content, just as beliefs have, and we can therefore have (primary) reasons for them.[9] Actions do not have propositional content, however. We are said to have reasons for them only because they are causal processes whose initial stages (so-called basic actions) are directly controlled by states for which we can have reasons. Therefore, the situation with respect to reasons for action is

---

[7] 'The Rationality of Belief and Some Other Propositional Attitudes', *Philosophical Studies*, 110 (2002), pp. 163-86, at pp. 174-5. Kelly's own favoured account is that reasons which I call primary are reasons for an attitude on which the attitude could be said to be *based*. I do not dissent; I have rather tried to explicate the basing-relation in terms of inferential links.

[8] Cf. T. M. Scanlon's view that what we have reasons for is 'judgment-sensitive attitudes' (*What We Owe to Each Other*, Harvard U. P., 1998, pp. 20-1). John Broome takes 'reason for/to' or 'ought' 'to govern a proposition' ('Normative Requirements', *Ratio*, 12 (1999), pp. 398-419, at p. 399). 'It is probable', 'it is necessary' and so on are operators which govern propositions. But when it is probable or necessary that p, what you have a reason for, ought or are required to, is *to believe* p, not p or that you believe p, whatever that may mean.

[9] More precisely, my view is that desires and emotions have propositional content in virtue of incorporating beliefs; see my *The Retreat of Reason — A Dilemma in the Philosophy of Life*, Oxford: Clarendon Press, 2005, pt. I.

perfectly in line with my claim that the fact that it is valuable to have an attitude, like the desire for a saucer of mud, provides a reason for wanting to have the attitude, not for the attitude itself. For the fact that (the action consisting in) causing a bodily movement will have valuable consequences provides a reason for wanting, intending or trying to cause it.[10]

We must however not confuse the claim that reasons for action are reasons for wanting, intending or trying to act with the (false) claim that reasons for action are only *secondary* reasons for action and primary reasons for wanting, intending or trying to act. The situation here is not parallel to what it is when we have to look around for primary reasons for an attitude in order to be able to bring it about that we have the attitude (for these reasons) and, thus, fulfil our higher-order attitude to have it . For in contrast to attitudes, actions *can* be produced at will, without any appeal to further reasons (which would be primary with respect to actions). We do not *bring it about* that we act (for some reason) by wanting, intending or trying to act (likewise, we do not *bring it about* that we perform non-basic actions by performing basic actions). This control as regards actions is what makes it true that (primary) reasons for wanting, etc. to act at the same time are (primary) reasons to act.

It follows from this that Rabinowicz' and Rønnow-Rasmussen's diagnosis of Gregory Kavka's Toxin Puzzle[11] is erroneous. They represent the relation between will and action to be something like the relation between higher- and lower-order attitudes when they write:

> I have a reason to intend an action (drinking the toxine [sic]) because I would be rewarded for having that intention, but, on the other hand, I have no reason to act on that intention (pp. 412-3).

There is however no space for me to have any reasons to act on an intention I have over and above the reasons I have for having the intention. When I have an intention, I do not have to decide, i.e., form an intention, whether to act upon it. That would start an infinite regress. So, there are no reasons in the light of which I could make this (alleged) decision. The connection between an intention (assuming it to be the immediate determinant of action) and the action fulfilling the intention is not governed by reasons, i.e., it is not the connection between a piece of reasoning and its conclusion. Whether you succeed in acting out your intention, or are struck by paralysis, prevented by a more powerful agent,

---

[10] In other words, Rabinowicz and Rønnow-Rasmussen commit what Kelly calls 'the Consequentialist Mistake': 'Because the expected consequences of performing an action bear on the rationality of that action, there is a persistent tendency to mistakenly assume that the expected consequences of holding certain propositional attitudes bear on the question of the rationality of those attitudes' (op. cit., p. 182). I am not sure, however, to what extent Kelly would accept my explanation of why this mistake is a mistake.

[11] ' The Toxin Puzzle', *Analysis*, 43 (1983), pp. 33-6.

etc., is not a matter of your rationality. It is a matter of extra-rational causal conditions, like your muscle strength, the absence of external obstacles, etc. Thus, in the Toxin Puzzle it is not the case, as Rabinowicz and Rønnow-Rasmussen maintain, that, in contrast to the intention, 'the action is irrational' (p. 413): a paralysis which prevented you from acting out your intention by imbibing the toxin would not save you from any measure of irrationality.

In the Toxin case, you cannot now intend to imbibe the toxin tomorrow afternoon because you realize that, after the reward time tomorrow morning has been passed, you will have no reason to drink the toxin. You can now intend to do an action at a future time *t* only if you now expect that the intention you now form could persist until *t* (and that you then have the ability and opportunity to act upon it), but this condition is not met in the Toxin case.

However, you could now have the second-order attitude of wanting to have this intention because you now have ample reason to want to have this intention: the fabulous monetary reward (which you reasonably covet). In other words, you now have secondary, but not primary, reasons to intend to imbibe the toxin tomorrow afternoon. Perhaps there are things you can now do to enable you to acquire this intention, e.g., hire a hypnotist who could deceive you into thinking that the reward will be paid out only *after* you have swallowed the toxin. You could then intend to acquire this intention, by doing these things. If you will take the toxin if you acquire this intention (we may assume that you have the ability and opportunity to put the intention into practice), you could also now intend *to see to it or to make it the case* that you tomorrow afternoon (intentionally) take the toxin. But you cannot now intend to take the toxin at this time.

At the end of the paragraph in which they discuss the Toxin Puzzle, Rabinowicz and Rønnow-Rasmussen point out that in cases of the demon sort '[t]he reward is not tied to our wanting to have a desire', but that '[i]t is dependent on the desire itself' (p. 413). This is true: the reward of escaping the demon's severe punishment is tied to, is a consequence of, desiring a saucer of mud. But this observation does not support their claim that the reward is a reason for this desire. For when what is reason-giving is 'dependent on' the object of a desire, i.e., is a fact about the object, the reason is not a reason for the object (whatever that could mean), but for having the desire towards it. Therefore, if what is reason-giving 'is dependent on the desire', we should expect that the reason is a reason for something that takes this desire as its object, e.g., a desire to have the desire. Accordingly, in the Toxin Puzzle, in which the reward is dependent upon the presence of a certain intention, the reason generated is a reason to want to have that intention or something like it.

Object-given and State-given Reasons

It may be wondered how Parfit's distinction between *object-given* and *state-given* reasons, criticized by Rabinowicz and Rønnow-Rasmussen, is related to my distinction between primary and secondary

reasons. Object-given reasons for a desire, Parfit writes, 'are provided by facts about this desire's object',[12] while state-given reasons for a desire are provided by facts 'about our *having* this desire' (ib., p. 22). Parfit claims that 'state-given reasons to *have* some desire are better regarded as object-given reasons to *want* to have it, and to try to have it' (ib., p. 24). If that is right, state-given reasons would coincide with secondary reasons, while object-given reasons would coincide with primary reasons.

Against Parfit's distinction Rabinowicz and Rønnow-Rasmussen object that state-given reasons can be rephrased as object-given reasons. This is so because if a pro-attitude towards an object would have a property *P*, the object has the property of 'being such that a pro-attitude toward it would have the property *P*' (p. 406). This is undeniable. Since the relevant object is an object of the state consisting in the having of a pro-attitude, a (full) statement of a state-given reason for the attitude will have to refer to this object. This paves the way for a re-formulation of the state-given reason as an object-given reason in the way indicated.

However, Parfit never claims that his distinction could solve the WKR problem; as Rabinowicz and Rønnow-Rasmussen note (p. 405), he never addresses this problem. What Parfit does is rather to set forth his 'value-based' theory of reasons in the terms of an FA analysis. Thus, the reason-providing facts Parfit has in mind are facts about the object of desire *which make this object good* and facts about our having a desire *which make it good to have this desire*. Now, it certainly does not follow that if a pro-attitude towards an object has a property *P* which makes the having of it good, the object's having the property of 'being such that a pro-attitude toward it would have the property *P*' would make the object good.

But, on this interpretation, there is a different objection, namely that Parfit's distinction becomes question-begging if proposed as a solution to the WKR problem (which, as remarked, Parfit never intended). The WKR problem arises because sometimes when we are said to have a reason to have a pro-attitude towards an object O, it is not O which is good but rather the having of the pro-attitude towards O. We therefore need to know when the reasons make the former rather than the latter true. It is of course circular to reply to this question by saying that the existence of the reasons make it true that O is good when they are provided by facts about O which make O good. To avoid this circularity we must obviously delete the reference to goodness. The question is whether the pseudo-Parfitian distinction this deletion issues in can avoid Rabinowicz's and Rønnow-Rasmussen's objection.

It seems to me that it can. We could reply to their objection that even if it is true that state-given reasons can be rephrased as object-given reasons, *genuine* object-given reasons for an attitude are still distinguishable from 'nominally' object-given reasons (and state-given reasons) for it by the fact that they *can* be so phrased that they do *not* refer to (the state of having) the attitude. That is, the reference,

[12] 'Rationality and Reasons' in Dan Egonsson, Jonas Josefsson, Björn Petersson and Toni Rønnow-Rasmussen (eds.): *Exploring Practical Philosophy: From Action to Value,* Ashgate, 2001, p. 21.

if any, to the attitude is *eliminable* in the case of genuine object-given reasons. Clearly, the kind of object-property Rabinowicz and Rønnow-Rasmussen give us must include such a reference (to a pro-attitude that would have *P*), i.e., the reference is ineliminable. Thus, the suggestion is that we can replace pseudo-Parfitian distinction with a distinction between (object-given) reasons for an attitude which contain no ineliminable reference to (the state of) having the attitude and reasons which are state-given in the sense of containing an ineliminable reference to the having of the attitude.[13] We can then ask whether the second group co-extensive with secondary reasons.

I think the answer is negative because some reason-giving facts are as much facts about the object as about the state of having a pro-attitude. A case discussed by Rabinowicz and Rønnow-Rasmussen could illustrate the point (p. 419). Suppose you admire the demon for the reason that he severely punishes everyone who does not admire him. This statement attributes to you a reason for your admiration of the demon which contains an ineliminable reference to this attitude, admiration for the demon. Now, this statement may mean that you have made yourself admire the demon for the reason that you will be severely punished by him if you do not have this attitude. No worry, this is a state-given reason that is secondary.

But the statement may also mean that you admire the demon for a primary reason to this effect, that you find him admirable because he is such an effective punisher that he severely punishes everyone who fails to admire him. Certainly, we may judge this a bad reason for admiration, but it may still be *your* (primary) reason for admiring the demon and finding him admirable: we do not always adopt our attitudes for good (primary) reasons, i.e., what in our eyes justifies does not always in fact justify. The statement may even combine these two senses: it may be that you have made yourself admire the demon for the primary reason that (he is such an effective punisher that) he severely punishes everyone who does not admire him, and have made yourself have this primary reason-based attitude for the secondary reason that the demon will otherwise punish you (for, on pain of punishment, he might have required you to admire him not for any old primary reason, but for this particular reason). In general terms, the problem is that it seems that some reason-giving facts can be classified both as facts about the object (e.g., the demon who threatens to punish you) and about the relevant attitude (the admiration you must have on pain of punishment).

Therefore, we cannot usefully claim that there is a perfect correspondence between the (revised) pseudo-Parfitian distinction and the distinction between primary and secondary reasons here adumbrated. As a matter of fact, I do not think that we could hope to be able to classify reliably reasons as primary and secondary on the basis of any marker in their content, and it is not part of my account that there is such a content-based criterion. The absence of something of this sort may well

[13] Jonas Olson makes a similar proposal in 'Buck-passing and the Wrong Kind of Reasons', *Philosophical Quarterly*, 54 (2004), pp. 295-300, at p. 299, and Christian Piller in'Content-Related and Attitude-Related Preferences' (forthcoming).

wreck the rival account Rabinowicz and Rønnow-Rasmussen put forward (pp. 414-20), as they fear, but it does not cause the slightest difficulty for the account here propounded.[14]

[14] Many thanks to Sven Nyholm, Jonas Olson, Derek Parfit and Christian Piller for valuable comments.