Buck-Passing and Non-Instrumental Reasons: A Solution to the Wrong Kind of Reason Problem

Lars Samuelsson (lars.samuelsson@umu.se)

Abstract

The wrong kind of reason problem (WKRP) has been posed as a problem for the so called “buck-passing account of value” (BPA). Although WKRP has been discussed at great length, none of the suggested solutions has gained general acceptance, and it seems a common view that the problem is yet to be solved or that it cannot be solved. In this paper I offer a solution to WKRP which draws on the notion of a non-instrumental reason. I suggest that only reasons that are in a certain sense non-instrumental are reasons of the right kind, and that the reasons involved in the standard wrong kind of reason cases are all what I call “purely instrumental reasons”.

Keywords

buck-passing, the buck-passing account of value, the wrong kind of reason problem, wrong kind of reasons, value analysis

1. Introduction

The wrong kind of reason problem (WKRP) is a much discussed problem for the so called “buck-passing account of value” (BPA). According to BPA, “being good, or valuable, is not

---

1 The phrase “wrong kind of reason problem” (as applied to BPA) was coined by Wlodek Rabinowicz and Toni Rønnow-Rasmussen, ‘The Strike of the Demon: On Fitting Pro-Attitudes and Value’, *Ethics* 114 (2004), p. 393.
a property that itself provides a reason to respond to a thing in certain ways. Rather, to be
good or valuable is to have other properties that constitute such reasons.” That is to say, BPA
identifies the property of being valuable with the property of possessing some other, lower-
order property (or set of properties) that provides reason for some relevant response (i.e.
positive responses for positive value and negative responses for negative value). Now, WKRP arises because there are cases – let us call them “WKR-cases” – where objects that
clearly lack value yet seem to have such reason-providing properties. Consequently, in order
to defend BPA against this problem one has to come up with a convincing explanation of why
the reasons in these cases are of the wrong kind, in the sense that the properties which provide
them do not give rise to corresponding values of the objects which possess them. More
precisely, what is needed in order to rebut WKRP is (a) a satisfactory way of discerning the
reasons that are of the wrong kind, and (b) a plausible explanation of why such reasons are of
the wrong kind.

However, this kind of problem had been acknowledged before with respect to relevantly similar views on value.

For instance, Sven Danielsson and Jonas Olson note that a version of WKRP was offered already by G.E. Moore
in his review of Brentano’s The Origin of Our Knowledge of Right and Wrong. See ‘Brentano and the Buck-

2 T.M. Scanlon, What We Owe to Each Other (Cambridge, MA: Harvard University Press, 1998), p. 97.

Although it was Scanlon who introduced the term “buck-passing” for this account, he is not its founder. It has
been traced back to e.g. Franz Brentano and A.C. Ewing. See Rabinowicz and Rønnow-Rasmussen, ‘Strike of
the Demon’, pp. 394-400; and Danielsson and Olson, ‘Brentano and the Buck-Passers’, p. 511.

3 I will not here get into the question of which these relevant (positive and negative) responses are.

4 Satisfactory in the sense that it (1) gives the right answer in every WKR-case, and (2) does not wrongly deem
any reason of the right kind a reason of the wrong kind.

5 Alternatively, one may argue that these reasons are not really reasons for the relevant responses at all, but
merely reasons for wanting to have these responses. See e.g. Derek Parfit, ‘Rationality and Reasons’, in D.
Egonsson, J. Josefsson, B. Petersson and T. Rønnow-Rasmussen (eds.), Exploring Practical Philosophy: From
Although WKRP has been discussed at great length, none of the suggested solutions has gained general acceptance, and it seems a common view that the problem is yet to be solved or that it cannot be solved. Rather than repeating these different suggestions I will here present a solution which seems to me capable of neutralizing WKRP by satisfying both (a) and (b). To anticipate, I will claim that only reasons that are in a certain sense non-instrumental are reasons of the right kind, and that the reasons involved in the standard WKRP-cases are all what I call “purely instrumental reasons”. The most crucial task in what follows


7 Perhaps needless to say, my solution bears similarities to (and is inspired by) some earlier proposed solutions – in particular, I think, those offered by, respectively, Stratton-Lake, in ‘How to Deal with Evil Demons’, and Lang, in ‘The Right Kind of Solution’ – but my hope is that it evades the various problems pertaining to these respective suggestions.
will be to characterize these purely instrumental reasons. For this purpose I will make use of Derek Parfit’s distinction between state-given and object-given reasons. This distinction has already been invoked in suggestions for solution to WKRP, but not in the way that I intend to use it. In contrast to these other suggestions, my solution does not rely on this distinction (i.e., the solution can be stated without invoking this distinction), but I find it helpful for the task of presenting and arguing for the solution that I propose.

A consequence of my solution is that mere instrumental value evades BPA. By saying that a thing has mere instrumental value I mean that it is (as far as this “value” is concerned) merely conducive to something else that is valuable, and that its “value” is merely reflective of the value of the thing to which it is conducive. Such “value”, I take it, is not really value at all; it does not represent a species of goodness. Things that are instrumentally valuable in this sense do not merit any responses in their own right. Some arbitrary brick in some wall is instrumentally valuable because it contributes to holding up the wall – which, we assume, is a good thing – but one may reasonably hold that the brick does not merit any positive responses of the kind that one takes to be relevant to BPA. If so, it does not – on this view – possess a value of its own, but is merely conducive to something else that is valuable. Hence its “value” should not be encompassed by BPA. Other instrumentally valuable things (i.e. things that are valuable on account of their property of being conducive to something else that is valuable)


9 I take this to be a common view. For instance, it seems to have been held by both Moore and Ross (G.E. Moore, Principia Ethica (Cambridge: Cambridge University Press, [1903] 1922), pp. 21-22; W.D. Ross, The Right and the Good (Oxford: Oxford University Press, [1930] 2002), p. 133), and it is explicitly endorsed by several contemporary moral philosophers (e.g. Jonathan Dancy, ‘Should We Pass the Buck?’, in A. O’Hear (ed.), Philosophy, the Good, the True and the Beautiful (Cambridge: Cambridge University Press, 2000), p. 159; see also Toni Rønnow-Rasmussen, ‘Instrumental Values – Strong and Weak’, Ethical Theory and Moral Practice 5 (2002), pp. 23-43). I will not here defend it any further.
may merit certain responses of the relevant kind. Perhaps we have reason to like the sun because of its indispensability to us. If so, the value of the sun is encompassed by BPA. This is all as it should be, I believe. To have genuine value is to merit responses in one’s own right. What that means more exactly will be clear once I have stated my solution to WKRP.

Before I develop this solution in detail (in sections three and four), we need a more thorough account of WKRP as well as of the state-given/object-given-distinction. The next section is devoted to this task. I end the article with two problematizing sections, one dealing with a possible objection to my solution, and one considering a kind of WKR-case that differs from the standard versions and might seem to require a different kind of reply.

2. The Wrong Kind of Reason Problem and State-Given Reasons

As WKRP was presented by Wlodek Rabinowicz and Toni Rønnow-Rasmussen, it “involves setting up a case in which no final value accrues to an object but there still are reasons to favour that object for its own sake, because such a pro-attitude has certain beneficial effects.”\(^\text{10}\) One of their (much discussed) cases involves an evil demon who threatens to inflict severe pain on us unless we admire him for his own sake.\(^\text{11}\) The demon’s threat seems to give us a reason to admire him, but it does not make him valuable (admirable). This case is a modification of an example formulated by Roger Crisp, where the demon will inflict severe pain on us unless we desire a saucer of mud.\(^\text{12}\) This fact about the saucer seems to give us a reason to desire it, but it does not make it desirable. Following Philip Stratton-Lake, let us call this example “ED1”, and Rabinowicz and Rønnow-Rasmussen’s modified version “ED2”.\(^\text{13}\)

\(^{10}\) Rabinowicz and Rønnow-Rasmussen, ‘Strike of the Demon’, p. 403.
\(^{13}\) Stratton-Lake, ‘How to Deal with Evil Demons’, p. 789.
Generally, what is common to the different versions of WKRP is that they aim to identify situations in which some object that lacks value yet has some property that provides a reason to respond to it in a positive way.\textsuperscript{14} Any case which succeeds in identifying such a situation appears to be a counter-example to BPA in Scanlon’s original formulation, since – on that formulation – the property of being valuable is none other than the property of possessing such reason-providing properties.

Having introduced WKRP Rabinowicz and Rønnow-Rasmussen begin to consider possible solutions. The first candidate they consider is one based on Parfit’s distinction between object-given and state-given reasons. Parfit writes: “Of our reasons to have some desire, some are provided by facts about this desire’s object. These reasons we can call object-given … Other reasons to want some thing are provided by facts, not about what we want, but about having this desire. These reasons we can call state-given.”\textsuperscript{15} On the face of it, our reason to desire the saucer of mud (in ED1) is not provided by a fact about the saucer, but a fact about the response – desiring – that we have reason to have towards it: Having this response shields us from pain, and that is our reason for having it. A natural suggestion, then, is that only object-given reasons are reasons of the right kind. But Rabinowicz and Rønnow-Rasmussen quite quickly dismiss this Parfit-inspired suggestion:

The distinction depends on a separation between facts about the object and facts about the attitude, or – to put it differently – it depends on a separation between the properties of the object and the properties of the attitude. It is only the former that can provide object-given reasons. However, drawing such a distinction does not yet solve our problem. It is easy to see

\textsuperscript{14} Actually, what is needed in order to raise WKRP is a case where there is reason for responses of the relevant kind towards some object X, but where the property (or set of properties) of X that provides this reason does not give rise to any value of X. It is not essential that X lacks value altogether.

\textsuperscript{15} Parfit, ‘Rationality and Reasons’, p. 21.
that for any property $P$ of the attitude there is a corresponding property $P'$ of the object: If a pro-
attitude toward an object $a$ would have a property $P$, then, ipso facto, $a$ has (or would have, if it
existed) the property $P'$ of being such that a pro-attitude toward it would have the property $P$. Consequently, to the attitude-given reason, provided by $P$, corresponds the object-given reason, which is provided by $P'$.

However, not everyone has found this dissolution of the state-given/object-given
distinction convincing. One who has raised doubts is Stratton-Lake. As he points out, there is
only one reason to desire the saucer of mud. It is not as if by re-describing the reason to desire
the saucer of mud as a reason provided by the saucer’s property of being such that desiring it
would shield us from pain we have thereby identified an additional reason to desire it (making
our overall reason to desire it stronger), besides the reason provided by the property of having
the desire (i.e. its property of shielding us from pain). I take this point – which will prove
important later on – to be uncontroversially valid. Thus it may seem as if we have to decide
which of the two properties – the property of having the desire or the property of the saucer of
mud – that really provides our reason to desire the saucer. And if so, it certainly seems more
natural to think of the reason as provided by the property of the state of desiring.

Stratton-Lake is of course right that we are dealing with two different properties here,
one of the state of desiring and one of the saucer of mud. And he is also right that there is only
one reason involved in ED1. However, a reason to respond to a certain thing is not itself a
property of that thing, but a fact about it. So the relevant question here is not whether we are

18 I take it that the participants in the debate about BPA and WKRP all agree that a reason (i.e. a normative
reason, which is the kind of reason involved here) is a fact that counts in favour of some response, or possibly a
property of such a fact, namely the property of counting in favour of. We may also note here that my proposal is
dealing with two different properties, but whether we are dealing with two different facts (of which only one can be our reason). And it seems to me that we are not. Rather, we are dealing with one fact that can be expressed in two different ways, either in terms of a property of the state of desiring, or in terms of a property of the saucer of mud (and which is a fact both about the state of desiring and about the saucer of mud):

\[(ED1\text{-state-given})\] The state of our desiring the saucer of mud has the property of shielding us from severe pain.

\[(ED1\text{-object-given})\] The saucer of mud has the property of being such that our desiring it shields us from severe pain.

If it is correct that these two sentences express the same fact, Stratton-Lake’s worry that we are positing one reason too many can be dismissed. In the words of Gerald Lang: “The truth, rather, is that the reason can be given by either type of property. These properties stand or fall together, in the sense that the reason for admiring the saucer of mud can be described either by invoking the properties of the attitudes, or the properties of the object.”

Even if it may seem more natural to think of the reason-providing property in ED1 as a property of the state of desiring the saucer of mud, this does nothing to show that it is in some sense incorrect to think of it as a property of the saucer of mud. Indeed, as Rabinowicz and Rønnow-Rasmussen explain in the quote above, the existence of a reason-providing property of the state of desiring the saucer of mud seems to imply the existence of a corresponding

---

reason-providing property of the saucer of mud. Thus it may seem as if the distinction between object-given and state-given reasons collapses and that we consequently cannot use it for the purpose of solving WKRP.

Now, I think this conclusion is somewhat premature, because the distinction can still be used to pick out two different sorts of reason, namely those reasons that can be described as state-given, and those that cannot. For while it is true that every state-given reason can also be described as an object-given reason, the reverse does not hold. It is not true of every object-given reason that it can be described as a state-given reason – that is only the case if the property of the object that provides the reason somehow has to do with our responses. So if we have reason to admire an ecosystem on account of its immense complexity, for instance, there is no way to re-describe that reason as a state-given reason. This fact makes the distinction a useful device, I think, for the purpose of providing a solution to WKRP.

The remainder of this paper is devoted to developing a solution to WKRP that makes use of the distinction between state-given and object-given reasons while admitting that it dissolves in the way described by Rabinowicz and Rønnow-Rasmussen. An important difference between my solution and the earlier suggestions built on this distinction is that I merely use it as a way of detecting the wrong kind of reasons – I do not claim that state-given reasons constitute a separate group of reasons. Differently put, the distinction is not necessary for my solution, but I find it helpful for the task of presenting and arguing for it.

3. Reasons as Opposed to Explanations of Reasons

As a first step towards developing my solution to WKRP, let me bring up yet another consideration from Rabinowicz and Rønnow-Rasmussen: “A property of an object that consists in it being such that a pro-attitude toward it would have the property \(P\) appears to be
quite artificial and somewhat unreal. It is a typical ‘Cambridge property’.²⁰ Perhaps, then, we can rule out – from being of the right kind – all and only reasons provided by such “Cambridge properties”. However, Rabinowicz and Rønnow-Rasmussen dismiss this option, because they think there are reasons of the wrong kind that are not provided by such properties. In particular, they take this to be the case in ED2:

The demon’s disposition to punish us if we don’t comply provides a strong reason for admiring him, which is of the wrong kind: it does not make the demon admirable in any way. But there is nothing Cambridge-like in this feature of the demon. If anything, it is more robust than the corresponding property of the attitude.²¹

Notwithstanding the point about robustness, the first claim in this quote seems to me mistaken. The fact that the demon possesses this property (the disposition to punish us unless we admire him for his own sake) does not really in itself provide any reason to admire him. In order to see why, let us remind ourselves of Stratton-Lake’s point that there is only one reason to admire the demon, namely that doing so shields us from severe pain. We have the following reason:

*(ED2-state-given)* The state of our admiring the demon for his own sake has the property of shielding us from severe pain.

As explained above, this (same) reason can also be expressed as object-given:

(ED2-object-given) The demon has the property of being such that our admiring him for his own sake shields us from severe pain.

Now, this fact about the demon is clearly different from the fact involving the robust property mentioned by Rabinowicz and Rønnow-Rasmussen:

(ED2-robust) The demon has the property of being disposed to inflict severe pain on us unless we admire him for his own sake.

An easy way to see that ED2-object-given and ED2-robust express different facts is to note that ED2-object-given may obtain without ED2-robust obtaining. ED2-object-given would still be true if, for instance, it was the demon’s sister (and not the demon himself) who had threatened to punish us unless we admire him for his own sake. Hence, unless we think there really are two different reasons to admire the demon in this case we have to disqualify ED2-robust from being a true reason to admire him.22

Somewhat peculiarly (given his insight that there is only one reason to admire the demon), Stratton-Lake writes as if there were in fact two different reasons involved here. He

22 Surely, we may (and often do) speak as if facts like ED2-robust were reasons. For instance, we may say that the fact that a certain dog – let us call him Boris – is angry, is a reason to avoid him. But strictly speaking this is not our reason. Our reason is provided (roughly) by the risk of being harmed by Boris. That Boris is angry explains why there is such a risk; it is not itself a reason in a strict sense (if it were a reason we would now have stated two different reasons to avoid Boris, which I take it that we have not). However, since the connection between angry dogs and the risk of being harmed is so obvious to us we can safely state the fact that Boris is angry as our reason to avoid him. (Loosely speaking, we can say that it is a reason, since it is a fact that counts in favour of avoiding Boris, but it is only an indirect, or derivative reason, which gains all of its reason-giving force from the risk of being harmed.)
states: “… a reason to have the pro-attitude toward the demon seems intuitively to be given by a feature of the demon…”\textsuperscript{23} and then, in a related footnote:

This is not to deny that if admiring the demon will avoid extreme pain, then there will also be a state-given reason to admire the demon. For my having this pro-attitude toward him will avoid extreme pain, and this fact about my admiration gives me reason to have this attitude. So the buck-passing account will imply that my admiration has instrumental value. This additional reason does not, therefore, cause a problem for BPV.\textsuperscript{24}

Now, I think this description of our reasons in this situation is misleading. The only reason (for admiring the demon) at work here is the reason related to the instrumental “value” of my admiration (ultimately grounded in the state of pain that I have a non-instrumental reason to avoid). There is no additional reason provided by the demon’s threat (or his disposition to punish us if we don’t comply). What sort of reason would that be? Stratton-Lake claims that it would not be an instrumental reason, since the demon’s threat is not a means to anything good. But it would, he writes, be a derivative reason, and that is what is supposed to make it a reason of the wrong kind:

But although his threat provides a noninstrumental reason to admire the demon, this reason is of the wrong kind to ground final value because it provides only a derivative reason. The fact that he has threatened me with severe pain unless I admire him provides me with a reason to admire him only because extreme suffering is something I have reason to avoid … Since the reason

\textsuperscript{23} Stratton-Lake, ‘How to Deal with Evil Demons’, p. 795.

\textsuperscript{24} Stratton-Lake, ‘How to Deal with Evil Demons’, p. 795 (emphasis mine).
provided by his threat is derived from the reason provided by the nature of suffering, the reason provided by his threat is derivative.\textsuperscript{25}

But the state-given reason, provided by the property of my admiring the demon, is of course also derived from the nature of suffering. Thus, on Stratton-Lake’s account, it seems as if the nature of suffering gives rise to \textit{two} different reasons to admire the demon in this case.

Now, in line with Stratton-Lake’s earlier remark, I do not see why we should believe there to be two different reasons at play here. So let me instead give an alternative picture of the situation. Rather than providing some additional derivative reason, the property of the demon \textit{explains} why there is a reason to admire him. The reason-structure in this case seems to me quite straightforward:

We have a reason to avoid pain (provided by the nature of the state of being in pain).

Thus we have reason to take the appropriate means to avoid pain.

Admiring the demon is an appropriate means to avoid pain (and that is our reason to admire him (cf. \textit{ED2-state-given})).

\textsuperscript{25} Stratton-Lake, ‘How to Deal with Evil Demons’, p. 797. As I noted in footnote 22, we may, when speaking loosely, refer to such facts as (derivative) reasons. But in strict theoretical contexts such as this one I prefer not to call them reasons at all, since referring to them as reasons may lead us to falsely believe that we have to explain them away as reasons of the wrong kind (which we do not, since, strictly speaking, they are not reasons to start with, but merely facts that explain reasons).
The fact that the demon is disposed to inflict severe pain on us unless we admire him for his own sake (i.e. \textit{ED2-robust}) explains why admiring the demon is a means to avoid pain (it is not itself a reason).

Now, where does this picture leave us? A first thing to note is that we have not yet been presented with an object-given reason of the wrong kind which is \textit{not} provided by a Cambridge-property (where that Cambridge-property is arrived at by way of re-describing, as an object-given reason, what is naturally thought of as a state-given reason), so perhaps all reasons of the wrong kind are such Cambridge-property-reasons after all. However, in order to make this suggestion work as a solution to WKRP we would have to be able to give a sufficiently clear characterisation of a “Cambridge-property”, and I would not want to rely on that. It would also have to be the case that (a) no Cambridge-property, characterized in the appropriate way, could ever provide a reason of the right kind, and (b) no property which is not a Cambridge-property could ever provide a reason of the wrong kind (unless it could be ruled out in some other way). And I would not want rely on that either.

4. Instrumental and Non-Instrumental Reasons

Another possibility – hinted at above – is that it suffices that a reason \textit{can be expressed as} a state-given reason for it to be of the wrong kind. That would correctly place both the reason to admire the demon and the reason to desire the saucer of mud among the reasons of the wrong kind, since both these reasons (as we have seen) can be expressed as state-given. I think this suggestion points in the right direction, but it cannot by itself do all the work that is needed. The problem is that some reasons that one may (not unreasonably) take to be of the right kind
can also be expressed as state-given. Consider the following example borrowed from Rabinowicz and Rønnow-Rasmussen:

(Lovable-object-given) Person P has the property of being disposed to respond with love to love.

As Rabinowicz and Rønnow-Rasmussen point out, “[w]e might well think this character trait makes him [P] lovable (with ‘lovability’ understood as a value term).” That is to say, it does not seem unreasonable to consider Lovable-object-given a reason of the right kind for loving P. But this reason can also be expressed as state-given:

(Lovable-state-given) Loving P has the property of being such that P is disposed to respond to it with love.

Hence, if we were to rule out all reasons that can be expressed as state-given from being reasons of the right kind, we would be ruling out too much; there are (or may be) reasons of the right kind that can be expressed as state-given. Thus, for a solution along these lines to work, we need a way to distinguish the state-given reasons (i.e. reasons that can be expressed as state-given) that are of the right kind from those that are of the wrong kind. The question we have to ask is what distinguishes the kind of state-given reason in Lovable-state-given from the kind of state-given reason in ED1-state-given and ED2-state-given. The answer I propose is that the reasons in the ED-cases are purely instrumental reasons.

26 Here it may be worth emphasizing that BPA is supposed to be a strictly formal account of value. It should be able to cover any fairly reasonable substantive view about which properties give rise to values.

In the ED-cases the reasons to respond to the object are grounded entirely in the (desirable) consequences of having these responses, whereas in *Lovable*, the idea is that loving is the correct, adequate, or fitting response to have towards a person who has the property of responding to love with love, irrespective of any consequences of having that response. Hence, the reason to respond in *Lovable* is not purely instrumental. As I intend to use the phrase “purely instrumental state-give reason”, a reason for a certain response R towards a certain object X is expressed as a purely instrumental state-given reason when it fits the following general pattern:

The state of having response R towards X has the property of being conducive to (or sufficiently likely to be conducive to) consequence C.28

As explained above, we can express what is naturally thought of as a state-given reason also as an object-given reason. Thus we have the following general pattern for a purely instrumental object-given reason, corresponding to the general pattern for a purely instrumental state-given reason:

X has the property of being such that having R towards X is conducive to (or sufficiently likely to be conducive to) C.

---

28 What counts as “sufficiently likely” depends on the circumstances. Sometimes we think there is reason to do something for the purpose of bringing about a certain (desirable) consequence even if the chance that our action (or response) will bring about this consequence is very small. On other occasions we want to be almost certain that our action will lead to a desirable consequence before we take there to be a reason to perform it.
My suggestion is that the reasons of the wrong kind in the ED-cases (and in other WKR-cases of that type) can be expressed as purely instrumental state-given reasons, while reasons of the right kind cannot.

I think it should be quite obvious why reasons that can take this form are of the wrong kind for corresponding to a value of X. These facts reveal nothing substantial about the object X. There could be all kinds of explanations of why having R towards X is conducive to the consequence in question, and depending on which of these possible explanations that is correct we will get different results with respect to the value of X. For instance, in making a verdict regarding the value of the demon it is relevant whether it is the demon himself or his sister who has threatened to punish us, and also what the explanation of this threat is – perhaps the demon has threatened to punish his sister unless she threatens to punish us; if so, that would arguably be relevant for deciding the value of the demon.

It is clear that the ED-cases involve reasons of the wrong kind according to this suggestion, since both ED1-state-given and ED2-state-given are examples of reasons expressed in the form of purely instrumental state-given reasons. Both these reasons are provided by the desirable consequence that we avoid severe pain. But is not Lovable-state-given also an example of a reason that may be expressed as a purely instrumental state-given reason, in which case my suggested solution would wrongly dismiss it as a reason of the wrong kind? Actually, it is not. Let us first note that, as I stated this reason above, it does not have the form of a purely instrumental state-given reason. Like ED2-robust it is expressed in terms of a disposition of the object in question (in this case P), and not in terms of any (desirable) consequence. But perhaps it can be re-described in the form of a purely instrumental state-given reason? To begin with, that would require the assumption that being

---

29 And, as argued above, ED2-robust is not, strictly speaking, a reason at all.
loved by P is good, so that the consequence that one is loved by P really does provide an instrumental reason for loving P. But granting this we get the following instrumental reason:

\[(Lovable-state-given-Purely-instrumental)\text{ Loving P has the property of being conducive to (or sufficiently likely to be conducive to) the consequence that P loves us.}\]

But just as \textit{ED2-state-given} and \textit{ED2-object-given} do not state the same fact as \textit{ED2-robust}, this sentence does not state the same fact as \textit{Lovable-object-given} and \textit{Lovable-state-given}. These latter sentences express a claim about a disposition of P, whereas \textit{Lovable-state-given-Purely-instrumental} expresses a claim about a consequence of responding in a certain way towards P. True, the disposition of P mentioned in \textit{Lovable-object-given} and \textit{Lovable-state-given} is such that the fact that P has it implies that responding to P with love is likely to have this consequence, but this does not mean that these sentences express the same fact as \textit{Lovable-state-given-Purely-instrumental}. That would be the case only if by saying that P has a disposition to respond with love to love, we mean nothing more than that loving P has the property of being conducive to (or sufficiently likely to be conducive to) the consequence that P loves us (and this is not what I take Rabinowicz and Rønnow-Rasmussen to mean, since they describe P’s disposition as a character trait, which I take to be something more “robust”). But that fact can hardly be a reason of the right kind to respond to P with love. That is to say, this can hardly be the kind of property that could give rise to a genuine value of P (i.e., understood this way, \textit{Lovable} is no longer a plausible example of a genuine value). I will return to this point in the next section.

Let us call any reason that can be expressed as a purely instrumental state-given reason a \textit{purely instrumental reason}, and any reason that is not a purely instrumental reason a \textit{non-}
*instrumental reason.* In contrast to the evil demon cases, there are indeed two different reasons in *Lovable* (given the two assumptions that being loved by P is good and that P really is lovable on account of his property of being disposed to respond to love with love): one that is purely instrumental (*Lovable-state-given-Purely-instrumental*) – corresponding to the instrumental “value” of loving P – and one that is non-instrumental (*Lovable-state-given/Lovable-object-given*) – corresponding to P’s lovability (i.e. his value). The first reason is of the wrong kind (in relation to P), whereas the second one is of the right kind (in relation to P). Hence, the suggestion under consideration gives the right answer in *Lovable* and similar cases. Using the distinction between purely instrumental reasons and non-instrumental reasons we can now formulate a version of BPA which I take to be immune to the ED-cases (and to other WKR-cases of the same type):

(BPA*) The property of an object X of being valuable is identical with its property of possessing some other (lower-order) property (or set of properties) p such that the fact that X has p is a non-instrumental reason to respond (in a relevant way) to X.

This version of BPA gives the right answer in the demon-cases and it allows us to consider P in *Lovable* valuable (loveable) on account of his disposition to respond to love with love.

---

30 To call such reasons non-instrumental is slightly misleading since there is a sense in which reason-providing properties such as possessing the disposition to respond to love with love can be said to be instrumental. After all, what is lovable about P’s disposition has to do with P’s treatment of others – those who love him – and the consequences that this treatment has for these persons (that they are loved and any further consequences associated with this consequence). This is why I have used the phrase “purely instrumental” instead of “instrumental”. What I mean by “non-instrumental” is thus really “not purely instrumental”. But for reasons of exposition I use the simpler expression “non-instrumental” when referring to reasons that are not purely instrumental.
5. A Possible Objection

So far so good. But suppose now that someone suggests that *Lovable-object-given-Purely-instrumental* actually is a non-instrumental reason to love P (or to adopt some other relevant positive response towards P), i.e. that P’s very property of being such that our loving him leads to his loving us grounds a genuine value (i.e. not only a mere instrumental “value”) of P. If we think that such properties might ground genuine value, then cases involving such reason-providing properties constitute counter-examples to BPA*. However, I think we should resist this suggestion. While BPA is supposed to be a formal account of value, and hence should be able to accommodate any fairly reasonable view as regards which properties ground value, it cannot be expected to accommodate unreasonable views. And I believe that the suggestion under consideration presents such an unreasonable view. The kind of property that figures in purely instrumental reasons as they are understood here does not seem to be the kind of property upon which a genuine value could supervene. Let me elaborate this point.

When a certain reason, r, to have a certain response, R, to a certain thing, X, is the fact that doing so is conducive to (or sufficiently likely to be conducive to) a certain consequence, C, there is obviously something else than X – something associated with C – that gives *this particular reason* (r) its reason-giving force. When we act on r by having R, we act for the purpose of bringing about C (or something associated with C). That is to say, there has to be something else than X that has some property the possession of which provides a reason to bring about C. It is this “something else” that has the genuine (positive or negative) value. In the evil demon cases, it is *being in pain* that possesses negative value (or that would possess negative value if it occurred). We have reason to avoid pain because of the nature of *being in pain*, the very same nature upon which its negative value supervenes.
Here we may also recall a point made earlier, that when a thing X has the property of being such that having R towards it is conducive to (or sufficiently likely to be conducive to) some consequence, there may be several different possible explanations of this fact. Depending on which of these explanations is correct, X may have either positive value, negative value or neither positive nor negative value (associated with this property). If the demon’s sister threatens to inflict severe pain on us unless we admire the demon, because the demon has threatened her with severe pain unless she does, one may reasonably hold that her property of threatening us is not associated with any properties of her that make her positively or negatively valuable. But if she threatens us because she thinks it is funny, her threat seems to be associated with properties that make her in some sense bad. Thus knowing merely that an object has this kind of purely instrumental property does not seem sufficient to make a well-founded judgment about its value. But if so, then arguably it cannot be the case that its value could supervene on this kind of property. If an object X with this kind of property has genuine value that is associated with that property, then this value has to supervene on some other property of X that explains why having R towards it has the consequence in question.

At this point another risk with my solution may have occurred to the reader. If mere instrumental “value” should not be encompassed by BPA, then it must be the case that we never have reasons of the right kind to respond to things that only possess mere instrumental value. Is that really the case? I believe the answer is “yes”. Consider a typical merely instrumentally valuable object, like a hammer (and let us assume that it is completely devoid of genuine value). What responses could we have reason to have towards such an object? We could have reasons for several different responses, such as taking good care of it, using it in certain ways but not in others, and even having certain attitudes towards it (such as some attitude of care). But as long as the hammer only has mere instrumental value there must always be something else of value whose properties ultimately explain these reasons and give
them their reason-giving force. So if I have reason to take good care of my hammer this is because doing so is conducive to (or sufficiently likely to be conducive to) some consequence that I have (or may come to have) reason to bring about (and that the hammer can help me bring about). It may be that some project that I have reason to bring about (or may come to have reason to bring about) is likely to require a good hammer, in which case my taking good care of my hammer is sufficiently likely to be conducive to the good consequence that this project is brought about. But this reason is purely instrumental and can be expressed as a purely instrumental state-given reason: Taking good care of my hammer has the property of being conducive to (or sufficiently likely to be conducive to) the consequence that my project is brought about. And that is my reason to take good care of my hammer.31

Now, in order to take good care of my hammer it may be required that I actually adopt an attitude of care towards it. If so, I clearly have a reason to adopt an attitude of care towards it. But, like the reason to take good care of my hammer, this reason is purely instrumental and can be expressed as a purely instrumental state-given reason: Adopting an attitude of care towards my hammer has the property of being conducive to (or sufficiently likely to be conducive to) the consequence that my project is brought about. And that is my reason to adopt an attitude of care towards my hammer. I cannot see how there could be a reason to respond towards an object with only mere instrumental value that could not be expressed in

31 Just to be clear, one may consistently hold that the hammer has genuine value (even final value) on account of its instrumental properties (see Christine M. Korsgaard, ‘Two Distinctions in Goodness’, *The Philosophical Review* 92 (1983), pp. 169-95; Shelly Kagan, ‘Rethinking Intrinsic Value’, *Journal of Ethics* 2 (1998), pp. 277-97; and Wlodek Rabinowicz and Toni Rønnow-Rasmussen, ‘A Distinction in Value: Intrinsic and For Its Own Sake’, *Proceedings of the Aristotelian Society* 100 (1999), pp. 33-51). But if so, the reason corresponding to this value cannot be expressed as a purely instrumental state-given reason (it has to be more like the reason in Lovable-state-given/Lovable-object-given). In my example with the hammer it is assumed that it only has mere instrumental “value”.
this form, and consequently I do not think that we have to worry about there being non-instrumental reasons to respond to such things.

6. Concluding Remark

I wrote in the introduction that what is needed in order to rebut WKRP is (a) a satisfactory way of discerning the reasons that are of the wrong kind, and (b) a plausible explanation of why such reasons are of the wrong kind. I believe that my solution to WKRP – expressed through BPA* – provides that: (a) Reasons that can be expressed as purely instrumental state-given reasons are reasons of the wrong kind. (b) Such reasons are of the wrong kind because the kind of property of an object X that could give rise to such a reason (i.e., the property of being such that having response R towards X is conducive to [or sufficiently likely to be conducive to] some consequence C) (1) reveals nothing substantial about X, and (relatedly) (2) is not the kind of property upon which a (genuine) value of X could supervene.

However, there is a kind of WKR-case that I have not yet brought up in this paper, which does not seem to share the pattern of the ED-cases. This kind of case has been most clearly stated by Crisp.32 Suppose I have promised to φ. This seems to give me a reason to choose to φ (where we suppose that choosing is a positive response of the relevant kind to ground value). But it does not seem to imply that φ:ing is therefore valuable, so this reason must be of the wrong kind. There is not space to discuss this kind of WKR-case at any length here, but let me mention what I take to be two potentially promising ways to deal with it:

1. I have previously suggested that the reasons in these cases do indeed depend on the consequences of responding to them, albeit conceptual consequences and not causal

---

consequences.\footnote{Samuelsson, ‘The Right Version’, pp. 389-92.} If that is correct, we can express the reason in Crisp’s example as a purely instrumental state-given reason as follows:

The state of choosing to \( \varphi \) has the property of being conducive to the (conceptual) consequence that I keep my promise. (And that is my reason to choose to \( \varphi \).)\footnote{For further discussion of this suggestion I direct the reader to Samuelsson, ‘The Right Version’, pp. 389-92.}

(2) When introducing BPA, Scanlon makes a distinction between \textit{valuing} something and \textit{claiming that} something \textit{is valuable}. He writes: “To claim that something is valuable (or that it is ‘of value’) is to claim that others also have reason to value it, as you do… claiming that something is valuable involves claiming that its attributes merit being valued generally…”\footnote{Scanlon, \textit{What We Owe}, p. 95.}

But the reason in Crisp’s example is strictly agent-relative; it is only I who have this reason to choose to \( \varphi \) (since it is only I who have made \textit{this} promise – others may also have promised to \( \varphi \), but they have not made \textit{my} promise to \( \varphi \)).

Philip Stratton-Lake and Brad Hooker make a similar point when they discuss a more general objection to BPA raised by Jonathan Dancy.\footnote{Philip Stratton-Lake and Brad Hooker, ‘Scanlon versus Moore on Goodness’, in T. Horgan and M. Timmons (eds.), \textit{Metaethics after Moore} (Oxford: Clarendon Press, 2006), p. 160.} According to Dancy, BPA has difficulties accommodating certain versions of deontology, since such theories may insist that there are reasons that are not related to any values.\footnote{Jonathan Dancy, ‘Should We Pass the Buck?’, pp. 168-70.} Thus, if I have promised to \( \varphi \) I have a reason to \( \varphi \), but there need not be any value involved. However, as Stratton-Lake and Hooker point out: “The buck-passing account of goodness can allow that an agent can have reason to
do some act which is in no way good. It would allow this if the features that gave the agent a reason to care about his doing act A did not give anyone else reason to care about his doing act A.\textsuperscript{38} Such a reason would then be a reason of the wrong kind.

\textsuperscript{38} Stratton-Lake and Hooker, ‘Scanlon versus Moore’, p. 160. Of course, one may think that there is a (non-instrumental) reason for any agent to admire instances of promise-keeping, but then one also has to think that these instances of promise-keeping (such as φ:ing in Crisp’s example) are valuable (since to be admirable is a way of being valuable). Cf. Stratton-Lake and Hooker, ‘Scanlon versus Moore’, pp. 160-61: “… if others have reason to approve of your doing A then your doing A must be good in some way.”