Against the 'First' Views

Why None of Reasons, Fittingness, or Values are First

Andrew Reisner

Against the 'First' Views

Why None of Reasons, Fittingness, or Values are First

Andrew Reisner

0. Introduction and Overview

Toni Rønnow-Rasmussen has been one of philosophy's most important contributors to our understanding of the nature of value, not least of all with respect to questions of whether different types of value are reducible to each other and whether value in general is basic. In this paper I make my own modest effort to follow in Toni's long shadow. The aim of this paper is to argue that there are at least two categories of normative or 'non-descriptive'[1] properties (in the terminology used in this paper) that cannot be reduced to other more basic non-descriptive properties and that one of those categories is that of value properties. Although the emphasis in the paper is on reductionist views, most of the arguments work equally well against a weaker category of view about the relations amongst non-descriptive properties, namely those that require different categories of non-descriptive properties to be linked by a necessary bi-conditional.

The arguments in this paper are incomplete in at least one rather obvious way. There are a number of candidates for basic non-descriptive properties,[2] but here I focus only on the three non-descriptive properties that receive the most attention in

---

[1] I recognise the problem with this terminology, insofar as 'non-descriptive' suggests that I am taking a linguistic or metaphysical stance on the nature of normative properties broadly construed, whereas I do not mean to do so. There are problems with other alternatives. Using 'normative' in this context makes it difficult to distinguish between the kinds of properties which are normative in a stricter sense, like oughts and reasons, and those which are not, such as evaluative properties. That same problem arises for using 'evaluative' to describe the category. I apologise to the reader for not finding a better term to use.

[2] I refer the reader to chapter 1 of Nils Sylvan's (2021) recent doctoral thesis for an excellent catalogue of candidate properties.

383

the literature: reasons, fittingness, and value. It is my conjecture, one for which I have no general argument at present, that with some work the arguments concerning the relations amongst those three properties can be adapted for use against reductionist programmes employing other combinations of non-descriptive properties. With that limitation in mind, I shall argue that no reductions are possible amongst these properties and thus that views that fall under the heading of 'reasons-first', 'fittingness-first', and 'value-first' – views that hold that there is a single most basic non-descriptive property – are false.

The arguments in this paper are directed in the main at metaphysical reductions, where one property is reduced to one or several more basic properties. There are at least two ways to argue against putative reductions of this kind. One is to show that the *analysandum* and the *analysans* have different necessary extensions. In work on value, this strategy is perhaps most familiar in the form of the *wrong kind of reasons problem* (WKR) for the fitting-attitude analysis of value.[3] WKR arguments are intended to show that there are instances in which there is a reason to favour x when x is not valuable. Likewise, there is the less commonly discussed *wrong kind of value problem* (WKV),[4] which aims to show the reverse, namely that there are instances where x is valuable, but where there is no reason to favour x or it is not fitting to favour x.

A second way to argue against attempted reductions is to show that despite the necessary extensional adequacy of the proposed analysis, the *analysans* lacks essential characteristics possessed by the *analysandum*, or alternatively adds features that in the relevant sense cannot be part of the *analysandum*. Needless to say, these two strategies do not exhaust the possibilities for arguing that an attempted reduction fails, but they are the two approaches that will be used in this text. These approaches differ in force in one important respect. Necessary extensional inadequacy is not only sufficient for showing that a reduction fails, but it is also sufficient to defeat a weaker claim, namely that there is a necessary bi-conditional equivalence between two or more categories of properties. The first strategy may thus be used to show that the *correctness conditions*[5] for one class of non-descriptive property cannot be given in terms of another, insofar as they are not necessarily extensionally equivalent. The second strategy does not show this directly. As at least some *-first* authors take *-first* claims to be about correctness conditions, they are only committed to necessary bi-conditional equivalence.

---

[3] It is undoubtedly fitting in this context to note that this problem was given life by the two classic Rabinowicz and Rønnow-Rasmussen (2004 & 2006) papers. The literature on this topic is now extensive. For some important examples, see Danielsson and Olson (2007), Lang (2008), and Olson (2009).

[4] See Bykvist (2009 & 2015), Dancy (2000), Heathwood (2008), Hurtig (2019), and Reisner (2015).

[5] Correctness conditions in this sense give the criteria for the conditions under which an object has a particular property. Understood this way, the fitting attitude analysis of good would say that an object is good only and always under the condition that it is fitting to favour that object without positing that what it is for the object to be good is for it to be fitting to favour.

Because of this, the extensional arguments tell against a wider understanding of what *-first* views amount to being.

In this text, I shall offer what I take to be decisive examples showing that neither fittingness nor reasons is necessarily extensionally equivalent to value, which suffices to show that an overarching *-first* project that aims to reduce two of reasons, fittingness, and value to the third property must fail. However, I shall also argue, using the second strategy, that reasons cannot be reduced to fittingness or to value, which tells us that at minimum reasons[6] and value are not less fundamental than fittingness.

1. Unalike Variance Conditions for Reasons and Value

Reasons and value have unalike variance conditions, or so I shall argue. And if they have unalike variance conditions, then that is enough to show, assuming that reasons and value are two of the three candidate non-descriptive property categories, that *-first* theories are false.

Any *-first* view with the ambition of being an analysis or a reduction must be built on a core bi-conditional that contains one of the non-descriptive properties on the lefthand side and another non-descriptive property of a different kind on the righthand side. These bi-conditionals are in general stronger than simple bi-conditionals, for example they may include determination and must in any case be necessary to play a role in an analysis. But since the present concern is with extensional inadequacy (from under-generation), it will suffice to work with simple bi-conditionals; if the relevant simple bi-conditional is false, then *a fortiori* so is a strengthened bi-conditional. Let us begin by focusing on the reasons version of the fitting-attitude analysis:

> 2. *The reasons version of the fitting-attitude analysis of value (RFAV)*:
> *x* is valuable if and only if there is a reason to favour *x*.

> F2. RFAV: *x* is valuable ⟷ there is a reason to favour *x*.

The target is to develop a schema for creating examples in which *x* is valuable, but there is no reason to favour *x*. One may start by considering the structure of reason relations:

> 3. *The simple reason relation*: Fact *f* is a reason for agent *A* to *ψ* to degree d[7]

---

[6] I suspect that in the final reckoning, one may need to treat oughts as irreducible to reasons. See Gjelsvik (2020) for a defense of the view that reasons and oughts cannot be reduced to each other.

[7] Some contemporary writers omit the final place in this relation. John Skorupski (2002, 2010) was careful to avoid this mistake in his pioneering work on the metaphysics of reasons.

In the simple reason relation, '[f]act *f* should be interpreted liberally so as to include conjunctions of facts or sets of facts.[8] The schematic variable ψ simply stands for anything for which there can be a reason (i.e. an action, belief, emotion, pro-attitude, etc). Crucially, reasons are indexed to agents.

1.1 The Under-generation Argument for Reasons and Value

With the essentials of the reason relation and RFAV having been set out, it is now possible to develop a schema for creating cases in which the lefthand side of the bi-conditional is true but the righthand side is false, thus showing that an analysis of value in terms of reasons to favour under-generates.

The simplest structure for such examples relies on descriptive, or if one prefers, non-normative *entanglement*.[9] One needs to generate examples in which favouring *x* makes *x* not be valuable.[10] I shall focus for now on *good* as a paradigm type of value. Here is a generic counterexample:

> 4. *The generic counterexample*: *x* is valuable at t_1 if and only if nobody ever has, does, or will favour *x*.

It is not difficult to fill out the details of this schema by making an appeal to sufficiently knowledgeable and powerful agents. Imagine that the demiurge has created a powerful entity whose nature is such that she relieves pain and suffering around the world anytime she waves her left arm, so long as nobody ever has, does, or will favour her waving her left arm. Her nature is also such that if anyone ever has, does, or will favour her waving her left arm, the effect of her doing so will instead be that she causes pain and suffering around the world. One may treat the effect of her waving her arm in both circumstances as necessary[11] due to her nature.[12]

---

[8] In Skorupski's (2002) explication of the reason relation, *f* stands for a set of facts.

[9] See Reisner (2015) and Risberg (2018) for detailed discussions of entanglement. The 'descriptive' qualifier is important; as Haim Gaifman argued as far back as the 1983, normative entanglement is highly problematic. I take this observation from Wlodek Rabinowicz's opposition at Olle Risberg's disputation.

[10] Strict covariance is also sufficient.

[11] An anonymous reader in another context pointed out to me that if one accepts S5, then this example is impossible, unless the entity in question exists. Given that philosophers often rely on possibly (but not in fact) necessary examples, one will have to take one's chosen solution to understanding this and other examples of this kind. Nothing about the example itself hinges on accepting S5. Only accepting K is required. I thank Jonathan Shaheen for a valuable discussion about this worry.

[12] In the past (2009 and 2015) I have treated FA as concerning final value. I assumed, too, that the final value of an action was the value of its consequences. I shall dispense with that assumption here for reasons that will soon be apparent.

An example of this form entangles favouring x (descriptive) with x's value (non-descriptive), or lack thereof. One can construct other such examples, of course, based on the same schema. Implicit in using an example of this form is the assumption that there is no reason to favour x if x will be bad, should one favour it. This underlying assumption seems highly plausible to me on its face.[13] Favouring x effaces the reasons for favouring x and thus defeats even the weakest guidingness constraints on reasons.[14]

Now we are in a position to see why value and reasons may have unalike variance conditions in the arm-waving example. So long as nobody ever favours the powerful entity's waving her left arm, it is good (valuable) that she waves her left arm. If somebody ever favours her waving her left arm, then it is bad (has disvalue) that she waves her left arm. Thus, whether or not somebody favours her waving her left arm changes the value valance of her waving her left arm. If we accept the argument about self-effacing reasons not being reasons at all, then there is never a reason to favour her waving her left arm. While the value valence of her waving her arm changes depending on whether or not anyone favours it, the valence of the reason to favour (i.e. a reason not to favour) never changes. And thus we have under-generation.

More needs to be said about this example, as I have as yet not specified what sort of value is at stake.[15] I shall consider three possibilities: intrinsic final value, extrinsic final value, and instrumental value.[16] It is at best unclear whether the entity's waving her left arm has intrinsic final value. The act itself, at least under that description, appears to be neutral. Perhaps the case could be reconfigured such that it has intrinsic final value, but I am unsure, so I shall assume for the moment that it does not. A second possibility is that the case has extrinsic final value. This seems more plausible to me. One might hold the view, for example, that the final value of an action is a function of the amount and distribution of wellbeing of its consequences.[17]

With respect to this case and others structured like it, whether something is extrinsically finally valuable will depend first on whether there is in fact such a thing as extrinsic final value and then on how one divides up the value bearers and background conditions. So perhaps the arm-waving example concerns extrinsic final value. It should be much less controversial to say that the arm-waving example

---

[13] I argue for this claim in §2.1.

[14] See Risberg (2020) and Rosenqvist (2020) for further discussion on guidingness. As Bruno Guindon pointed out to me, guidingness constraints are often understood in some sort of deliberative internalist terms, i.e. that one can do what there is a reason to do by including the reason in one's deliberation. The guidingness constraints that are relevant here are extremely weak and fully consistent with rejecting all forms of deliberative constraints.

[15] The importance of clarifying what sort of value applies in this example was pointed out to me by Antti Kauppinen, who also provided advice I have followed here in structuring the discussion.

[16] I have left out a discussion of possible differences between *value for someone* and *value simpliciter*. For a discussion of the latter in the context of FA, see Rønnow-Rasmussen (2007, 2011 & 2021).

[17] This is perhaps John Broome's (2004) view in *Weighing Lives*.

is a case of instrumental value. The arm-waving case thus creates clear difficulties for a reason-to-favour analysis of instrumental value. It may create difficulties for an analysis of final value that includes extrinsic final value, and it does not yet pose a straightforward difficulty for analysing intrinsic final value.

A second example is required to create clear difficulties for an analysis of intrinsic final value.[18] Let us suppose, as many philosophers have, that it is intrinsically finally valuable to love another person unconditionally.

This example also involves a demiurge who decides this time that if anyone ever favours a particular instance of Xenophon's unconditionally loving any particular person, he will never unconditionally love that person. The demiurge's decision has the peculiar effect that it is impossible to favour a particular (actual) instance of Xenophon's unconditionally loving another person, because the existence of the pair {Xenophon loves x unconditionally at t_1, anybody ever favours that Xenophon loves x unconditionally at t_1} is impossible. No instance of Xenophon's loving another person can be favoured while there is a reason to favour it, because if it is favoured, there will be no such instance. Put another way, the demiurge's condition makes favouring particular (actual) instances of Xenophon's unconditionally loving another person metaphysically impossible.

One may find parallel cases when it comes to reasons for action and value. Suppose that one offers the following bi-conditional claim about beauty:

> 5. *The beauty bi-conditional*: x is beautiful if and only if there is a reason to have an aesthetic experience of x.

We should understand 'have an aesthetic experience of x' as encompassing actions such as viewing paintings, listening attentively to symphonies, watching films, etc. Now consider a delicate sandstone rock formation whose unique beauty can only be experienced from the changing perspectives given by climbing its face. Regrettably the rock is delicate enough that even the lightest touch of its surface destroys those natural features that make it beautiful, rendering its beauty impossible for anyone to experience.[19]

In this case, presumably the features that make the rock formation beautiful do so whether or not they can be experienced.[20] Thus so long as one does not climb the formation, it remains beautiful. But if one is climbing or has climbed the formation,

---

[18] This example was proposed to me by Jaakko Kuorikoski. I am grateful for his suggestion.

[19] Randall Harp expressed to me the worry that there are no beautiful objects that could only be experienced in this way, as perhaps an object that is beautiful, but that cannot be experienced, is not in fact beautiful. I do not share this intuition, but I have no argument against it that does not rely on one's already sharing my intuition that there are such objects. Bruno Guindon expressed concern that the example itself suggests the implausibility of the beauty bi-conditional.

[20] Objectivism of this sort about beauty is controversial. Nonetheless, I follow Elisabeth Schellekens (2006) in accepting an adequate degree of objectivity for the purposes of this example.

then the formation is not beautiful, due to the destructive effects of climbing it. There is no reason for one to experience the formation, because doing so effaces the physical features of the formation that provide reasons to experience it; one has no (aesthetic) reason to climb the formation once one is climbing it. Here again, we see that there is no reason for one to climb the formation, irrespective of whether one climbs it or not, but the formation is beautiful if one does not climb it and is not beautiful if one does.[21]

It bears noting at this point that although RFAV is formulated as a simple bi-conditional, the counterexamples would also hold for a counterfactual version of the principle. In all relevantly similar worlds, the same entanglements would exist.

1.2 Objections to the Counterexample Schema

It is of course fair to ask whether the assumption that self-effacing (putative) reasons to favour are not actual reasons to favour is correct. I believe it is, but I would like to look at two possible objections against the force of cases built on the entanglement schema.

The first objection posits that there is a reason for someone in another possible world to favour the entity's waving her left arm, since that person would sit outside the actual world's past, present, and future. I find this proposal very odd, but a parallel proposal has been suggested to me with respect to fittingness. There are a number of technical issues that arise with respect to this proposal, many of which I have discussed in depth in an earlier paper.[22] However, I am now convinced that there is a (somewhat) more straightforward way to reply to this objection than my previous attempt, at least with respect to reasons.

Note that this objection is describing a possible reason to favour the entity's waving her left arm, not an actual (in the modal sense) reason to favour it. This would mean that RFAV would have to be modified:

> 2a. *Possible reasons fitting-attitude analysis of value* (PRFAV): x is good in the actual world if and only if there is a possible reason to favour x's occurrence in the actual world.[23]

Although the arm-waving case is stated in general terms, it has specific implications. If it is generally good for the entity to wave her left arm, so long as it is never favoured, then each specific existentially quantifiable occurrence of her waving her

---

[21] Simon-Pierre Chevarie-Cossette suggested another example of a beautiful painting that blinds anybody who looks at it before they can see it.

[22] Reisner (2015).

[23] I have not noticed any commitments specifically to this view in writing. Despite that, it has often been suggested to me in correspondence and conversation as a way to solve the sorts of difficulties raised by WKV.

left arm (when nobody favours her doing so generally) is also good. A successful analysis of *good*, or of any sort of value, and the bi-conditional on which it is built, will entail that each specific instance of the entity's waving her left arm is good under the condition that (eternally) nobody favours it.

PRFAV implies that there is someone in another possible world who has a reason to favour one or more specific occurrences in the actual world in which the entity waves her left arm. This is because reasons are indexed to individuals, or sets of individuals. A reason needs to be a reason for at least one particular individual to be a reason at all. It is doubtful that individuals in other possible worlds can favour an entity in the actual (from our perspective) world's doing anything at all, because favouring that occurrence would require having that occurrence in mind. And it is itself doubtful that we can have singular thoughts about individuals or specific events in other possible worlds,[24] which is what would be required to get a particular individual (situated in a particular world) in mind. If nobody *can* have the reason, then nobody *does* have the reason; therefore, it is not the case that there is a reason for *x* to favour that such-and-such occurs in another possible world.

But suppose that it is possible to have singular thoughts about individuals or events in other possible worlds. In that case, PRFAV itself seems like a bad principle, in part because it would over-generate in a peculiar way.

Suppose that a powerful being will improve life in another possible world (which is not the actual world) each time someone in the actual world[25] performs a cruel act that causes only pain. Someone in that other world has a reason to favour the performance of those cruel acts in the actual world, namely that they reduce suffering in her world. According to PRFAV, the fact that she has a reason to favour their occurrence in the actual (from our perspective) world also makes them good in the actual world, when it instead is right to say that they are bad in the actual world, although their occurrence in the actual (from our perspective) world is good in her world.

Of course talk about what merely possible rather than actual individuals have reason to favour in the actual (from our perspective) world is strange in numerous ways, not least of all because it is difficult to understand the idea at all without accepting modal realism. Otherwise, it is not clear that there are in an interesting sense individuals in other possible worlds.[26] The very claim that *x* is good if a merely possible person favours it sounds false. Strangeness aside, PRFAV is extensionally inadequate, which is enough to reject it without complaining about the metaphysics.

---

[24] *Ibid.* and see Soames (2002).

[25] The *actual* operator indexes to this world, whereas 'another possible world' should be taken to indicate the use of a different indexical operator *W*, which functions like the *actual* operator but localises to the world in which it is being used. I discuss how this operator works in Reisner (2015).

[26] On a modal anti-realist view, one might wish to treat worlds as logically consistent complete state descriptions. Such descriptions would include descriptions or representations of individuals, but not actual individuals.

The second objection concerns the 'eternity' condition in the counterexample to RFAV, namely that it is implausible to say that the entity in the example's actions could be affected by what occurs in future, perhaps because of an assumption that the future is open and thus non-determinate. I do not have very much to say about this objection, because it clearly hinges on the difficult question of whether the future is determinate, or perhaps knowable. I suspect that if the future is non-determinate or non-knowable, complications will arise, too, for versions of RFAV that rely on the possibility or existence of reasons in future to favour the entity's waving her arm. I shall simply concede for the time being this remains an unaddressed potential objection.

2. The Argument Extended to Fittingness

If the argument in §1 is correct, then reasons-first is ruled out, because it is extensionally inadequate on any interpretation. This still leaves the possibility that a fittingness-first view is correct. In this section, I argue that fittingness-first is false, most importantly because the fitting-attitude analysis of value is extensionally inadequate, under-generating in some circumstances and perhaps over-generating in others.

However, I shall begin by looking at another potential problem, one astutely identified by Christopher Howard.[27] The problem is that fittingness on traditional views seems to under-generate with respect to reasons, at least if one accepts that there are state-given reasons for propositional attitudes. Howard's account is cleverly constructed so as to avoid cases in which fittingness under-generates with respect to reasons.

My presentation of Howard's view is not entirely faithful to the original, but the changes affect small details that are distracting to include in this context and not the central extensional adequacy concerns.[28] His account is built on two main claims:

> 6. *Value as fittingness (VAF)*: x is non-instrumentally good if and only if it is fitting to favour x.

And

> 7. *Reasons as fittingness (RAF)*: There is a reason to favour x if and only if: 1) it is fitting to favour x, or 2) it is fitting to favour that one favour x.

VAF is just FA. RAF, read with the first disjunct alone, says that there is a reason to favour x if and only if it is fitting to favour x. That would appear to rule out state-

given reasons. For example, it would be ruled out that one ought to desire to listen to Vogon poetry[29] to avoid being thrown off of a Vogon ship, although Vogon poetry itself lacks desirable qualities.[30] Intuitively, it is good to desire to listen to Vogon poetry, because it is good to avoid being cast out into the vacuum of space. According to VAF, it would follow that it is fitting to desire that one desires to listen to Vogon poetry. Howard stipulates that when a second-order desire is fitting, then there is a reason to have the first-order desire. This resolves the under-generation problem for state-given reasons.

However, notice that Howard's view still entails that x is good only if one has a reason to favour x. That is because the righthand side of VAF and the first disjunct on the righthand side of RAF specify the same condition, namely that it is fitting to favour x. Thus when it is fitting to favour x, x is good and there is a reason to favour x.

Yet this is problematic in light of the arguments in §1. They show that reasons under-generate with respect to value, i.e. that there are some cases in which x is good, but there is no reason to favour x. That conclusion is inconsistent with Howard's view:

1) x is good iff it is fitting to favour x (Ass. VAF)

2) If it is fitting to favour x, then there is a reason
     to favour x (Ass. sufficient cond. in RAF)

3) If x is good, then there is a reason to favour x (from 1, 2)

4) Not: If x is good, then there is a reason to favour x (Ass. from §1)

5) Conclusion: Either 1, 2, or 4 is false   (from 1-4)[31]

Assuming we do not reject premise 4, then this raises a problem for Howard's view: namely that either VAF is false or that RAF is false and consequently that all-in his view is false. If nothing else, this points to the difficulty of constructing a fittingness-first account that implies that there are state-given reasons for propositional attitudes.

Nonetheless, for now I want to focus on FA/VAF and show that it is false. To do so, I shall introduce a new version of WKV for fittingness. I shall take up the question of whether one of either reasons or fittingness might be first relative to the other in §3.

---

[29] For more on Vogon poetry, including some examples, see Adams (1981).

[30] For purposes of the example, I assume that Vogons can tell whether one has a desire to listen to their poetry or whether one is merely acting as though one does.

[31] I thank Jens Johansson for pointing out a problem, now remedied, with an earlier version of this argument.

2.1 Some New Arguments against the Fitting-attitude Analysis of Value

There are, as far as I can see, two strategies for showing that fittingness and value have unalike variance conditions. One strategy is the strict argumentative analogue of the arm-raising or unconditional love argument presented in §1 against RFAV. One need only swap in 'fittingness' for reasons and fix the grammar accordingly to see how such an argument would look.

However, there is a complication. The argument in §1 relied on adopting what I shall call the 'realisability condition for reasons' (RCR):

> 8. *Realisability condition for reasons (RCR)*: Fact *f* is a reason for agent *A* to ψ to degree *d* only if *A* can (metaphysically) ψ whilst there is (still) a reason for *A* to ψ to degree *d*.

As I noted when the idea was presented informally in §1, it is difficult to doubt this condition, which may be understood as an extremely weak guidingness constraint.[32] A parallel condition would be required to transfer the same argumentative structure to fittingness. That would give us a *realisability condition for fittingness (RCF)*:

> 9. *Realisability condition for fittingness (RCF)*: It is fitting for *S* to favour *A*'s ψ-ing only if *S* can favour *A*'s ψ-ing whilst it is (still) fitting for *S* to favour that *A* ψs.

Intuitions about this principle may be less clear than they are for RCR. However, I suspect that most people will find RCF difficult to doubt on reflection.

It may help to begin by thinking about fittingness outside the context of FA. Consider these fittingness claims, some with synonyms for 'fitting' to avoid leaning too heavily on a single word for evidence:

> F1: It is fitting to feel gratitude towards Sophia, but not if you feel gratitude towards her.

> F2: It is appropriate to be angry at Harvey, but not if you are or become angry at Harvey.

> F3: It is correct to hold your fork in your left hand, but not if you hold your fork in your left hand.

> F4: It is meet to honour Achilles, but not if you honour Achilles.[33]

Each of F1-F4 would be a pretty odd thing to say. Presumably, they are odd to say, because they each imply a conditional claim of the form: If you will feel/do *x*

---

[32] I thank Bruno Guindon for pointing out to me that I ought to say this explicitly.

[33] I thank Jimmy Goodrich for suggesting a valuable revision to these examples.

towards A, it will not be fitting/appropriate/correct/meet to feel/act that way. Or perhaps it implies a counterfactual version of the same claim. It would be bemusing, if not vexing, to be told that it is appropriate to hold one's fork in one's left hand, only then to be told that holding one's fork in one's left hand is inappropriate on account of the fact that one is holding one's fork in one's left hand. One might be forgiven for worrying that one has fallen through the looking glass. These examples are, of course, not dispositive. Perhaps the relevant intuitions rest on social factors that are not indicative of the nature of fittingness itself. However, they are at least suggestive.

Let me offer what may be a stronger consideration in favour of RCF. The entanglement cases I have been discussing are instances of the following general schema:

C1F: It is fitting that S favour A's ψ-ing only if S does not favour A's ψ-ing.[34]

Particular events can be fitting to favour, too:

C1Fp: It is fitting that S favour that instance of A's ψ-ing only if S does not favour that instance of A's ψ-ing.

If RCF is correct, then no cases for which either C1F or C1Fp are true. Conversely, if there are cases for which C1F or C1Fp are true, we must reject RCF. If one could find a reading of C1F or C1Fp where there were cases that seemed intuitively correct, then assuming that other reasonable conditions are met, we could reject RCF. As I shall argue briefly here, it is difficult to see what kind of reading would do the trick.

One way to try to find acceptable cases of C1F and C1Fp is to see if we can find a helpful interpretation of 'favouring A's ψ-ing'. Both the most natural reading and what strikes me as one promising-seeming alternative interpretation are problematic. One way to read the phrase is with a universal quantifier: all favourings of A's ψ-ing are fitting for S. But C1F and C1Fp entail that no favourings of A's ψ-ing are fitting. Consequently, this reading is simply false if there are any favourings of A's ψ-ing.

Another possible reading of C1F and C1Fp is that 'favouring A's ψ-ing' should be understood as expressing an event (or mental state) type. Since the existence of a type does not entail the existence of tokens of that type, it seems open in principle that it could be fitting for S to favour A's ψ-ing, *qua* type, without S's ever favouring a token instance of A's ψ-ing. This reading is better, but still problematic, because the type features in a relation in which none of its tokens can feature. Of course, there are some relations in which types can feature in which their tokens cannot due to category problems, e.g. those relations in which the relevant *relatum* must be an abstract object and the type's tokens are concrete objects.

---

[34] S and A need not be different individuals, but they of course may be.

In this case, however, it is difficult to see why the fittingness relation could not take an individual instance of favouring as a *relatum*. Thus the situation remains odd. Consider a parallel case. The type, Charles Maturin's *Melmoth the wanderer*, contains a greater number of nested narratives than either the type or a complete token of Edgar Allan Poe's 'The cask of amontillado'. It is impossible that a complete token of *Melmoth the wanderer* contains fewer nested narratives than either the type or a complete token of 'The cask of amontillado'.

It is generally, but not universally the case that tokens share the relevant properties of their types. Given that there is no difficulty with the existence of complete tokens of favouring event/state types, it seems to me that interpreting 'favouring A's ψ-ing' as being about an event or mental state type does not render C1F or C1Fp true, at least not without further argument. In order for the use of types to work, one would have to be happy with the existence of types with complete tokens that do not share in principle shareable properties and relations with the type itself, where the failure to share in those properties is not due to category problems.[35] To the best of my knowledge, there has been very little work done on spelling out the conditions under which tokens inherit properties or roles in relations from their types, and thus I make the foregoing comments with all due caution.

A final interpretation of C1F and C1Fp is that 'favouring A's ψ-ing' expresses an existentially quantified claim about actual or possible favourings. C1F and C1Fp remain false on this interpretation, as no actual or possible instances of S's favouring that A ψs make them come out as true. One can make the modal point explicit:

> C1F*: It is, or would be, fitting that S favour A's ψ-ing only if S does not, or would not, favour A's ψ-ing.

Someone who wishes to deny RCF must offer another interpretation of 'fitting to favour' that is consistent with the fact that there are no possible instances of favouring, actually or counterfactually, that have the property of being fitting.

Thus far I have been discussing these cases with the assumption that S and A are in the same world. As far as I can see, the remaining option is to allow that S and A exist in different worlds. I have already mentioned some difficulties with doing this,[36] but I shall set those aside. The arm-raising example poses no problem for FA, if we allow trans-world fittingness – its being fitting for an individual in one world to favour events or states-of-affairs in another – into the analysis.
However, trans-world fittingness has its own difficulties. In particular, it over-generates for value. I can offer two kinds of example of over-generation. The first is the example of attitudes that are fitting on comparative grounds:

---

[36] See §1 and Reisner (2015).

10. *Comparative admiration*: It is fitting to admire individuals, the moral character of whom is substantially higher than our own and than that of those around us.

In the actual world, this is at least a plausible fittingness principle. In a scene in the movie, *Rocky*, Rocky Balboa is watching a fight on TV at a local bar. Apollo Creed wins, but the bartender dismisses Creed as a chump. Rocky is appalled and criticises the bartender, saying that at least Creed took his best shot, remarking that the bartender has not done anything remotely so worthy with his life. Rocky is of course impressed that Creed won, but he also admires his dedication to developing his talents.[37] The admonition and the admiration would be out of place if Creed's efforts were merely typical of those made (up to that time) by Rocky himself, the bartender, and the other 'bums[38] from the neighbourhood', even if many other top boxers train equally as hard.

If we accept *comparative admiration*, or any other fittingness claim with a similar structure, we end up with the following problem. Suppose that *S* lives in a possible world occupied only by people of low moral character. *S* (somehow) comes to learn about *A*, who exists in a different possible world. Although *A* is in fact a pretty awful person by the standards of *A*'s world, he is a paragon of virtue compared to those who inhabit *S*'s world. It is fitting for *S* to favour *A*, but it is clearly not the case that *A* has the property of being admirable in *A*'s own world. That is the first example of over-generation.

Here is a second. If we accept the strange picture on which people in one world can get those in other worlds in mind, the following is a possible case. Individuals in *S*'s world take the greatest pleasure from the existence of feats of daring-do in other worlds. In her own (different possible) world, *A* sets out to climb its tallest mountain. It is fitting for *S* to favour that *A* climb the mountain, because *A*'s doing so is good in *S*'s world due to the pleasure that her doing so causes there. But let us suppose that *A*'s climbing the mountain in her own world will lead her to install the relay that will bring Skynet online. Her climbing the mountain is bad in her own world. It is fitting for *S* to favour that *A* climb the mountain, but it is not good in *A*'s world that she do so, violating the core bi-conditional of FA.[39]

Therefore, I conclude that value cannot be reduced to fittingness, and I have likewise argued that value cannot be reduced to reasons. This entails that value is

---

[37] For people concerned about *Rocky* interpretation, this point is made explicit in the temporally distant sequel, *Creed*.

[38] Henry Hill expresses a similar sentiment, although in his case about being a 'schnook', in *Goodfellas*.

[39] Peter Fritz pointed out to me the extreme bizarreness of the metaphysics required to make sense of this example, and I can only agree. However, it seems to me that someone who wished to use trans-world fittingness as a way to resolve the worries I have raised about FA would have to accept similarly bizarre metaphysics. I should certainly be content to see the entire approach of using trans-world fittingness ruled out as beyond the pale of reasonable metaphysics. I am regrettably not in a position to make that judgement or the required argument myself.

not subject to analysis or necessary bi-conditional equivalence in the manner required for fittingness-first and reasons-first theories. This rules out any understanding of -first views that entail at least as much a necessary bi-conditional equivalence.

3. Reasons and Fittingness

We are now left with a final question: is one of reasons or fittingness first relative to the other? I believe the answer to this question is 'no', but I have no conclusive argument to offer to that effect. Instead of offering a conclusive argument, I wish to turn to Danielsson & Olson's influential 2007 paper on FA.

When Danielsson & Olson set out to solve the wrong kind of reason problem, they did so by importing a non-descriptive notion, *correctness,* that appeared to be in some important way distinct from *being a reason.* Correctness is fittingness. Their strategy was initially to divide reasons into two kinds: those that arise directly from correctness ('content reasons') and those that do not ('holding reasons' that are not also content reasons). The former are suitable for FA, and the latter are not.

Importantly for the present discussion, Danielsson & Olson then pursue a reductive project in the later part of the paper, developing a Ewing-inspired account of how to reduce all holding reasons to content reasons. Because content reasons are nothing more than facts about its being correct or fitting to hold certain attitudes, Danielsson & Olson's project is in the final analysis an early version of fittingness-first.

We can see the same general idea if we look back to Howard's reasons-as-fittingness condition. He offers a way of accounting for non-correctness reasons in terms of fittingness. I have already introduced Howard's account in some detail, and it is worth considering again in this context.

According to Howard, there is a reason to have a pro-attitude with contents c if it is fitting to favour c or if it is fitting to favour favouring c. This second condition is perhaps necessarily co-extensional with Danielsson & Olson's holding-but-not-content reasons. Let us suppose that it is. A proposed advantage of Howard's view is that it offers conceptual gain.[40] But conceptual gain comes at the cost of theoretical unity. The relationship between fittingness and reasons looks *ad hoc*, with the second disjunct of the bi-conditional introduced only to ensure extensional adequacy (to preserve the existence of state-given reasons for propositional attitudes).

Perhaps one might want to defend the introduction of the second disjunct by pointing out that on Howard's view, this makes sense of reasons' being sensitive to (changes in) value. Reasons' sensitivity to value is explained by the underlying

---

[40] A term I borrow from Rabinowicz (2008 & 2012) to describe a reduction in the number of categories of concepts or properties in a particular (e.g. normative) domain.

relation between fittingness and value on the one hand and fittingness and reasons on the other. If it is fitting to favour *x*, then *x* is good, according to Howard. And if it is fitting to favour favouring *x*, then favouring *x* is good. Correspondingly, there is a reason to favour *x*, namely that *x* is good. And favouring *x* itself turns out to be good when there is a reason to favour favouring *x*.

However, if, as I have argued, there is no bi-conditional equivalence between its being fitting to favour *x* and *x*'s being good, then the relation between reasons and fittingness, if there is one, does nothing to explain whatever relation there is between reasons and value. The loss of theoretical unity and explanatory unity seems to sap the independent motivation for accepting reasons as fittingness, making it look like it is an *ad hoc* principle designed to ensure extensional adequacy alone.

To this end, I am more strongly inclined to think that a view like that offered by Conor McHugh and Jonathan Way[41] is well supported by considerations of theoretical unity, despite still being false. On their view, one has a reason to desire *x* only if it is fitting to desire *x*, excluding Howard's additional disjunct that there is a reason to desire *x* if there is a reason to desire to desire *x*. They stand with philosophers such as Derek Parfit and John Skorupski in suggesting that all reasons are reasons of the right kind for the fitting-attitude analysis.[42] And according to McHugh & Way, this fact is meant to be explained by the primacy of fittingness.

Whether one favours the Howard-style approach or the McHugh & Way-style approach to fittingness-first, there is a basic problem that neither account can avoid. Fittingness does not do the work of reasons. The central feature that underlies reasons, oughts, and other properly normative properties is that they are guiding in some loose sense. This sense is loose enough that it need not include any link between being (potentially) motivated by a consideration and that consideration being a reason, but not so loose that the realisability condition is violated. Note that I am not assuming that not violating the realisability is sufficient for possessing guidingness. This seems to put a fittingness-before-reasons view onto the horns of a dilemma. If fittingness is not a properly normative property, then there is more to something's being a reason than its being fitting: a new feature, guidingness, is added. On the other hand, if fittingness is as guiding as reasons, fittingness then looks rather like a normative property, perhaps so much so that one doubts that there is anything more to being fitting than being a reason that obtains in virtue of certain kinds of relations between an attitude and its contents. In that case, it is most natural to interpret reasons as being prior to fittingness, perhaps making fittingness reducible to reasons.

If the arguments in the rest of this chapter are correct, and fittingness is not prior to value, then there seems to be no special reason to believe that fittingness is in general more basic in the relevant sense than other non-descriptive properties.

---

[41] McHugh & Way (2016 & 2022).

[42] Parfit (2001) and Skorupski (2002 & 2010) take these reasons to be object given reasons. However, the spirit of their views and that of McHugh and Way are much the same.

This is clearly not a conclusive argument against the claim that fittingness is prior to reasons. However, properly normative notions are central to much of our ethical and even epistemological theorising, and if we are not willing to abandon the weak guidingness that I claim is the characteristic feature of the normative, then it is difficult to see how fittingness will in any interesting sense be prior to reasons. Perhaps the reverse is true as well, but I shall let the matter rest there.

4. Conclusion

In this paper I have argued that value is not analysable in terms of reasons or fittingness, due to the extensional inadequacy of such analyses. The fact that value over-generates for fittingness also means that fittingness cannot be reduced to value. This is sufficient to show that *-first* views that have the ambition to reduce two of fittingness, reasons, and value to the remaining third property category are false. I have not taken up the interesting question of the right aims or ambitions of *-first* projects. If the arguments here are correct, that is unnecessary. The least ambitious version of the *-first* projects is to provide adequacy conditions for all non-descriptive properties in terms of just one non-descriptive property, even when there are no analytic or reductive ambitions in play. Even this least ambitious project cannot survive the falsification of the relevant bi-conditional claims. More ambitious projects will necessarily imply more, and are *a fortiori* also false.

The arguments in §3 are incomplete, but perhaps suggestive of the claim that reasons cannot be analysed in terms of fittingness. Whether the reverse is true is uncertain, but I see no special grounds for optimism that such an analysis is possible.[43]

Works Cited

Adams, Douglas (1981). *The Hitchhiker's Guide to the Galaxy*. New York: Pocket Books.

Broome, John (2004). *Weighing Lives*. Oxford, Oxford: Oxford University Press.

Bykvist, Krister (2009). No Good Fit: Why the Fitting Attitude Analysis of Value Fails. *Mind* 118 (469): 1-30.

Bykvist, Krister (2015). Reply to Orsi. *Mind* 124 (496): 1201-1205.

Dancy, Jonathan. (2000). Should We Pass the Buck? *Royal Institute of Philosophy Supplement* 47: 159-173.

Danielsson, S. and Olson, J. (2007). Brentano and the Buck-Passers. *Mind* 116 (463): 511-22.

Gaifman, Haim. (1983). Paradoxes of Infinity and Self-Applications, I. *Erkenntnis* 20 (2):131-155.

Gjelsvik, Olav (2020). Reason and Oughts: Fundamentals with the Normative. *Acta Philosophica Fennica* 96: 155-177.

Heathwood, Christopher. (2008). Fitting Attitudes and Welfare. *Oxford Studies in Metaethics* 3: 47-73.

Howard, Christopher. (2019). The Fundamentality of Fit. *Oxford Studies in Metaethics* 14: 216-235.

Kiesewetter, B. (2018). Contrary-to-Duty Scenarios, Deontic Dilemmas, and Transmission Principles. *Ethics* 129 (1): 98-115.

Lang, Gerald. (2008). The Right Kind of Solution to the Wrong Kind of Reason Problem. *Utilitas*, 20 (4): 472-89.

McHugh, C. & Way, J. (2016). Fittingness First. *Ethics* 126 (3):575-606.

McHugh, C. & Way, J. (forthcoming). *Getting Things Right: Fittingness, Reasons, and Value*. Oxford: OUP.

Olson, Jonas (2009a). The Wrong Kind of Solution to the Wrong Kind of Reason Problem. *Utilitas* 21 (2):225-232.

Parfit, Derek (2001). Rationality and Reasons. In Egonsson, Dan, et al. (eds.) *Exploring Practical Philosophy: From Action to Values*. Aldershot: Ashgate: 17-39.

Rabinowicz, Wlodek (2008). Value Relations. *Theoria* 74 (1): 18-49.

Rabinowicz, Wlodek (2012). Value Relations Revisited. *Economics and Philosophy* 28 (2): 133-164.

Rabinowicz, W. and Rønnow-Rasmussen, T. (2004). The Strike of the Demon: On Fitting Pro-attitudes and Value. *Ethics*, 114 (3): 391-423.

Rabinowicz, W. & Rønnow-Rasmussen, T. (2006). Buck-passing and the right kind of reasons. *Philosophical Quarterly* 56 (222):114-120.

Reisner, Andrew (2009). Abandoning the Buck Passing Analysis of Final Value. *Ethical Theory and Moral Practice* 12 (4): 379-395.

Reisner, Andrew (2015). Fittingness, Value, and Trans-World Attitudes. *Philosophical Quarterly* 260: 1-22.

Risberg, Olle (2018). The Entanglement Problem and Idealization in Moral Philosophy. *Philosophical Quarterly* 68 (272): 542-559.

Risberg, Olle (2020). *Guiding Concepts: Essays on Normative Concepts, Knowledge, and Deliberation*. Thesis, Uppsala University.

Rosenqvist, Simon (2020). *Hedonic Act Utilitarianism: Action Guidance and Moral Intuitions*. Thesis, Uppsala University.

Rønnow-Rasmussen, Toni (2007). Analysing Personal Value. *The Journal of Ethics* 11 (4):405-435.

Rønnow-Rasmussen, Toni (2011). *Personal Value*. Oxford: Oxford University Press.

Rønnow-Rasmussen, Toni (2021). *The Value Gap*. Oxford: Oxford University Press.

Schellekens, E. (2006). Towards a Reasonable Objectivism for Aesthetic Judgements. *British Journal for Aesthetics* 46 (2): 163-177.

Skorupski, J.M. (2002). The Ontology of Reasons. *Topoi* 21 (1-2): 113-124.

Skorupski, J.M. (2010). *The Domain of Reasons*. Oxford: Oxford University Press.

Soames, Scott (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of" Naming and Necessity"*. Oxford: OUP.

Sylvan, Nils (2021). *Fittingness and Partiality: On the Partiality Problem for the Fitting Attitude Account of Value*. Doctoral thesis, Stockholm University.