I. NOT FITTINGNESS, NOT REASONS, AND NOT VALUE: AGAINST THE ‘FIRST’ VIEWS

[This is a draft chapter, chapter IV in the current plan, for my book project *The Pragmatic Foundations of Theoretical Reason*. It is still very much in progress. Comments are welcome. If you would like to cite the chapter in a general way, it is far enough along for that. If you want to cite some specific argument or claim, it would be best to check with me first in case I’ve made changes.]

O. INTRODUCTION

Future historians of philosophy looking at the current times may wonder about the causes of the current obsession with finding what normative or evaluative concept or property is first, and what the temptation was to pursue these projects with at least partially reductive ambitions within the normative or evaluative domain. Individual positions such as reasons first, fittingness first, and value first have all individually been criticised, often by proponents of one of the other -first views. This chapter has a broader aim and develops an argument aimed at showing that at least straightforward versions of each of these views are false.¹

---

¹ Perhaps a partial answer is the present state of the philosophy profession in which younger philosophers and emerging mid-career scholars feel pressured to be associated with clear views expressed with great confidence, irrespective of whether the degree of confidence is matched by the degree of support for the view.

² Several authors defend the fittingness-first approach. Some important examples include: Chappell (2012), McHugh & Way (2016), and Howard (2018). Reasons first perhaps the most prominent treatment in the literature. For book-length treatments of different versions of the view, see Lord (2018) and Skorupski (2010). For further discussion of the motivations behind reasons-first views, see Schroeder (forthcoming). Value-first is defended in Maguire (2106) and perhaps has antecedents in the work defending scalar utilitarianism, for example Norcross (2006).

³ ‘All’ is strictly an exaggeration. The arguments here do not directly touch on debates about ought + analyses of reasons, such as the view that reasons are explanations of oughts or that reasons are evidence of oughts. Although I believe these views to be false as well, the master argument in this chapter provides no support my belief.
More specifically, the central aim of this chapter is to establish the claim that at least two of the three non-descriptive notions used in various value-first views – reasons, fittingness, and value – have unalike variance conditions, and that value in particular has unalike variance conditions with both fittingness and reasons. What I mean by ‘unalike variance conditions’ is that sometimes one category of non-descriptive properties will change valence (from good to bad, or fitting to unfitting, for example) in circumstances in which another does not.

Showing that different non-descriptive concepts or properties have unalike variance conditions is one way to show that those concepts or properties are not identical, or that one is not analysable in terms of the other, due to their not being necessarily co-extensional, which is the minimum condition for an analysis, reduction, or an identity claim.⁴

I have argued elsewhere⁵ that fittingness and value vary in unalike ways under various pairs of circumstances. I concluded from that argument that the standard version of the fitting-attitude analysis of value was false. However, the original argument cannot easily be separated from more complicated questions about the nature of modality. Here I develop what I hope is a more straightforward argument that reasons and value have unalike variance conditions. After giving that argument, I offer a new, simpler argument that fittingness and value have unalike variance conditions. If these arguments are right, then value is one of at least two normative primitives.⁶

This much view leaves open the possibility that at least one of the non-descriptive properties may be reducible to one or a combination of the other two. I shall suggest that fittingness is the

---

⁴ See Block & Stalnaker (1999) for an influential defence of the view that necessary co-extension is sufficient for identity and Chalmers and Jackson (2001) for a defence of the view that it is not. There are other ways to argue against there being any correct value-first view. Daniel Wodak (forthcoming) argues from broader metaphysical consideration against the general value-first project in the non-descriptive domain.

⁵ See Reisner (2015).

⁶ It is important to distinguish this project from the work of authors like Maguire (2016) and his value-first programme, which I take to be a non-starter. It is a consequence of the arguments here that value first is false, because there will be at least one other ‘equal’ normative primitive.
most likely candidate for reduction or elimination.

1. An Overview

Moral philosophers are accustomed to working with two at least apparently distinct families of broadly non-descriptive concepts and, mutatis mutandis, properties: evaluative concepts and deontic concepts. Evaluative moral concepts may be thin, such as good or bad. Alternatively they may be thick, such as admirable or cruel. The deontic moral concepts, including right, wrong, and morally permitted are often understood as being special cases of non-moral deontic concepts, namely ought, forbidden, and permitted simpliciter.⁷ We may for convenience group the deontic concepts under the wider umbrella of normative⁸ concepts, which also includes normative reasons, but which does not include either evaluative concepts or correctness concepts such as fittingness.⁹

Some substantive moral theories are expressed in terms of relations between evaluative concepts or properties and normative concepts or properties. A simple form of utilitarianism may be formulated thusly: It is wrong to φ when there is at least one other available action, the consequences of which are better than those of φ-ing. Setting aside projects that have treated utilitarianism as giving a real definition of ought,¹⁰ in general work on consequentialism makes no special claim at all about whether normative concepts can be analysed in terms of evaluative ones or vice-versa. And likewise with respect to normative and evaluative properties. Nonetheless, ethics scholars of many different stripes are disquieted by the lingering possibility that at bottom there are unanalysable basic non-descriptive concepts or properties. While they may have to accept at

---

⁷ See Zimmerman (2015) for more discussion.
⁸ Note that here, as is the practice throughout this book, I am using ‘normative’ in a special sense and not as a blanket term for a broad class of non-descriptive discourse.
⁹ ‘Normative’ is often used to pick out the family of concepts and properties which I am calling ‘non-descriptive’. At times, ‘evaluative’ is used that way, too. But I mean to use the names for the concepts and properties in the matter I stipulate.
¹⁰ E.g. Moore (1903)
least one type of non-descriptive concept or property, they maintain a the-fewer-the-better approach.¹¹ For reasons that are not well-explained by the authors themselves, it seems that evaluative concepts and properties have been a particular source of anxiety.¹²

1.1 The fitting-attitude analysis of value

One strategy that has its contemporary roots in the work of Franz Brentano is to employ some version of the fitting-attitude analysis to give a conceptual and/or metaphysical analysis of good and other value notions in terms of fitting pro-attitudes.¹³ For the moment, let us focus on the metaphysics. The property of being fitting is sometimes understood as its own basic type of property and sometimes understood as being identical to the property there’s being a normative reason, or at least a certain type of normative reason. The so-called fitting-attitude analysis of value (FA) can be stated thus:

1. x is valuable if and only if x is the fitting object of a pro-attitude.

Or perhaps more popularly:

1a. x is valuable if and only if it is fitting to favour x.

In 1a) ‘favour’ is a generic pro-attitude verb. In this instance, the formulations should be read neutrally insofar as ‘fitting’ may be understood either as a distinct concept or property from that of there’s being a (certain type of) reason or as being identical to there’s being a (certain type of)

¹¹ Wlodek Rabinowicz calls this ‘conceptual gain’, see Rabinowicz (2008 & 2012).
¹² This point emerges clearly in Lang (2008).
¹³ See Danielsson & Olson (2007) for a more detailed discussion of the history of the fitting attitude analysis of value.
1.2 From the fitting-attitude analysis to first-ism more generally

After having largely faded from philosophical discussions following the 1960s, the fitting-attitude analysis was brought back into vogue due to the work of T.M. Scanlon, who proposed a closely related view in the form of the buck-passing account of good.¹⁴ The details of the buck-passing account need not detail us here. More importantly for present purposes, with the subsequent publication in 2004 of Włodek Rabinowicz and Toni Rønnow-Rasmussen’s paper, ‘The Strike of the Demon: On Fitting pro-Attitudes and Value’, it became common to accept a formulation like those in 1) and 1a) as opposed to the buck-passing formulation favoured by Scanlon.

Authors, including Rabinowicz and Rønnow-Rasmussen, saw significant potential in the fitting-attitude analysis to ease worries about explanatory and both metaphysical and conceptual complexity by giving a (presumably reductive) analysis of value concepts and/or properties in terms of fittingness (understood as reasons in Rabinowicz and Rønnow-Rasmussen’s work and later as a sui generis property by other authors).

It is important to highlight two of the purported advantages of the account. The first is what Włodek Rabinowicz calls ‘conceptual gain’.¹⁵ By reductively analysing one non-descriptive concept (or property) in terms of another, one decreases the number of basic non-descriptive concepts or properties. Rabinowicz’s interest in conceptual gain presaged the more comprehensive -first projects.

The second purported advantage is that relations amongst reasons or fittingness (according to the interpretation of ‘fitting’ in use) and value are, if not explained, at least accounted for.¹⁶ Value

---

¹⁴ Scanlon 1998. For a helpful discussion of what buck-passing really amounts to, see Olson (2009)
¹⁵ See fn. 12.
¹⁶ I discuss difficulties with this kind of explanation in Reisner (2009a).
and reasons, or fittingness, co-vary because value reduces to reasons or fittingness.

It suffices to note that authors who endorse the fitting-attitude analysis see the above as advantages of the view.¹⁷ It suffices, because it helps to explain how the -first views arise. These desiderata may in principle be used to motivate other -first projects, including those that seek to reduce fittingness to value or reasons. To the extent that one sees this pair of desiderata as motivating the fitting-attitude analysis, it is wholly natural to pursue an even more complete reductive project taking any one of the non-descriptive concepts or properties and reducing the other two to it.

And this is the essence of the -first projects. By picking a single non-descriptive notion to treat as fundamental, or first, one increases the conceptual or metaphysical gain and also extends the explanatory(-ish) project. And indeed, each of value, fittingness, and reasons have been proposed as the relevant first notion.¹⁸

The -first projects may have further ambitions. For example, they may seek to provide analyses of other apparently non-descriptive concepts or properties, such as rationality,¹⁹ or of apparently descriptive ones such as modality and probability.²⁰ The strategy of finding unalike variance conditions is fully generic, and perhaps it may be fruitfully applied to these and other further ambitions. I shall not investigate the matter further in this chapter.

2. Lessons from the fitting-attitude analysis

The central argument in this chapter is that variance conditions are the same for at most two of the three non-descriptive notions.²¹ The two notions with conditions that are most clearly unalike

¹⁷Ibid. One may indeed wonder, as I do, whether they really are advantages.
¹⁸Citations
¹⁹Lord (2020)
²⁰Skorupski (2010).
²¹I use ‘notions’ here when what the discussion applies to both concepts and their corresponding properties.
are value and reasons, or so I shall argue. If they can be shown to be unalike, then all three of the first programmes are unsuccessful, at least in their simple forms. I shall take up the task of arguing that value and reasons have unalike variance conditions in §3. In this section, however, I would like to revisit the longstanding debate about the fitting-attitude analysis of value, focusing primarily on the current debate about FA in which fittingness is understood as an unanalysable concept or property. I shall argue that FA fails in §4, but for the moment rest content to make some salient methodological observations that will provide context for the rest of the chapter. The discussion here only concerns properties, but both the discussion and the arguments it contains may for the most part be freely adapted to apply to concepts.

2.1 The abandonment of the wrong kind of reason problem

Before turning to fittingness as a sui generis property, I shall begin by recalling why fittingness came to be seen that way, as opposed to being another way to express the claim that there is a normative reason for something. Philosophers working on topics in normativity and value theory are by now familiar with discussions about the wrong kind of reason (WKR). Lost to some extent today is the original usage of this term, deriving from Rabinowicz and Rasumussen’s 2004 discussion22 of WKR.23 For WKR, the meaning of ‘wrong kind of reasons’ is ‘wrong kind of reason for use in the fitting-attitude analysis’, i.e. they are kinds of reasons that lead to over-generation of value when they appear in the analysans. I confess to not understanding how the concept of a reason’s being of the wrong kind applies in the other contexts in which it is now used,24 but the defects of the post ‘Strike’ literature need not detain us. The original invocation of the wrong kind

23 This regrettable loss is likely due to Pamela Hieronymi’s (2005) confusing use of the term.
24 For examples of the current usage, see Gertken & Kiesewetter (2017) and Sylvan & Lord (2019). In some of these papers, it is clear which reasons are reasons of the wrong kind, but it is unclear what the upshot is to being a reason of the wrong, or right, kind. I thank Daniel Star for pressing me to include this clarification.
of reason arises for an understanding of ‘fitting to favour’ as ‘there being a reason to favour’. On this interpretation, we can set out the reasons version of the fitting-attitude analysis of value:

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 \( x \).\(^{25}\)

Drawing on an example of Roger Crisp’s in which an evil demon pressures us to desire a bowl of mud despite the undesirability of the bowl of mud itself,\(^{26}\) Rabinowicz and Rønnow-Rasmussen note that the analysis, at least at first blush, appears to over-generate. Crisp’s bowl of mud example is intended to show that there is a reason to favour something that is not itself good or otherwise valuable. The easiest way to construct these examples is by using *state-given reasons*, i.e. reasons for holding propositional attitudes due to certain consequences of holding the attitude rather than due to the attitude’s relation to its contents.\(^{27}\) For example, intuitively there is no reason to desire to eat a bowl of mud simply for the sake of eating a bowl of mud. However, there is a reason to desire to eat a bowl of mud, if one would be severely punished for not desiring to do so.

As Rabinowicz & Rønnow-Rasmussen and Christian Pillar\(^{28}\) respectively observed, cases like Crisp’s bowl of mud example could be handled by RFAV, if one distinguished between the right and wrong kind of reasons for the fitting-attitude analysis. And designating state-given reasons as the wrong kind and object-given reasons (those that obtain in virtue of some relation between an

---

\(^{25}\) In order to achieve conceptual gain, RFAV must be read as an analysis, and thus ‘with determination going from right to left’ should be added. For present purposes it is enough to work with the simple bi-conditional.

\(^{26}\) Crisp (2000).

\(^{27}\) Although I and many other authors associate state-given reasons mainly with consequentialist considerations, D’Arms & Jacobson (2000) provide an early and influential example of a non-consequentialist state-given reason.

attitude and its contents)²⁹ as the right kind provides a neat solution to the simplest form of WKR.

However, what Rabinowicz and Rønnow-Rasmussen gave with one hand, they took away with another, developing the now well-known pairing problem,³⁰ which appears to resist the state-given reasons as wrong reasons and object-given reasons as right reasons solution to WKR.³¹ There has been much subsequent literature on WKR and possible solutions³² with inconclusive results.

Nonetheless, an important turning point came in 2007 with the publication of a paper by Sven Danielsson and Jonas Olson, ‘Brentano and the Buck-Passers’,³³ in which they propose to identify the right kind of reasons with what they call ‘correctness reasons’. A correctness reason is not itself a normative reason. Those are called ‘holding reasons’ by Danielsson & Olson, but correctness reasons, according to Danielsson & Olson, always give rise to holding reasons, although not all holding reasons have corresponding correctness reasons. Correctness reasons are facts that make the relation between an attitude and its content correct. For example, that $x$ has certain properties – that that make it desirable – is a correctness reason to desire $x$. That I would get a prize for desiring $x$ is a holding reason for desiring $x$, but not a correctness reason, as that does not make $x$ itself desirable.

In this picture, holding reasons that lack corresponding correctness reasons are reasons of the wrong kind, whilst those that have corresponding correctness reasons are reasons of the right kind for RFAV. The important innovation in Danielsson & Olson’s work is the use of the notion of correctness to solve WKR. And correctness is fittingness. Because its being correct (having a correctness reason) to favour $x$ is on Danielsson & Olson’s view not the same thing as there’s being

²⁹ See Danielsson & Olson (2007) and Reisner (2009a, 2014 & 2018) for more discussion.
³⁰ Rabinowicz & Rønnow-Rasmussen develop a generic schema to convert state-given reasons into object-given reasons, thus rendering the distinction unhelpful for solving WKR.
³¹ The pairing problem is the result of an argument demonstrating that there is a procedure generating an object-given reason out of each state-given reason. The new object-given reasons are not suitable for use in FA.
³² See Lang (2008), Olson (2009), and Samuelsson (2013) for further discussion.
³³ Danielsson & Olson (2007).
a reason (in the standard sense of the term) to favour x, although the former entails the latter, they achieved only partial conceptual gain. If we take them for the sake of argument to have solved WKR, then we still have a basic non-descriptive ontology with both fittingness and normative reasons as essential parts.

Despite the richer ontology, we now have a picture in which one no longer needs per se to discuss normative reasons directly and their relationship to value. Rather, it is enough that we can (let us suppose) create an extensionally adequate bi-conditional linking fittingness (now replacing ‘correctness reasons’) and value without reference to normative reasons. And thus we have the essential idea behind fittingness first account of FA without room for WKR.

2.2 A clue from the wrong kind of value problem

Assuming for the sake of argument that moving from normative reasons to fittingness suffices to resolve, or render obsolete, WKR, then there are two remaining ways to try to argue that FA is false. The first is to argue that it is a bad analysis for other reasons. This is an approach that Roger Crisp and I have each addressed in earlier work. Although I remain convinced that FA is a bad analysis for reasons independent of concerns about extensional adequacy, demonstrating that FA is extensionally inadequate remains the most definitive and secure method to show that it is false. Thus we may turn to the second alternative, which is to demonstrate that FA under-generates, i.e. that there are cases in which x is good but it is not fitting to favour x.

Two arguments have been developed that attempt to create what one might dub the ‘wrong kind of value problem’ (WKV). The first and most influential is due to Krister Bykvist and his solitary goods objection. Bykvist’s objection invites us to consider a possible world which contains a single all-things-considered pleasure-experiencing egret and no other individuals past, present, or

---

34 Crisp (2005) and Reisner (2009a)
35 See Bykvist (2009). Thoughtful criticisms of Bykvist’s view may be found in Orsi (2013).
future possessed of a mental life. According to hedonism, we should say that that world is a good (value) containing world, as it has an individual experiencing net pleasure and no individuals experiencing net pain.

By stipulation, there are no individuals in the egret’s own world for whom it is fitting to favour the egret’s being happy, or their world’s being such that the egret is happy. If we assume that there must be at least one historical individual for whom it is fitting to favour $x$ in order for it to in fact be fitting to favour $x$, then it is not fitting in the egret’s own world to favour the egret’s being happy.

And according to Bykvist, it would not be fitting for us, or any creatures, to favour the egret’s being happy. This is, very roughly, because Bykvist takes the important (for the fitting-attitude analysis of good) favouring attitude to be desire. Accordingly, it cannot be fitting for us to desire states-of-affairs in worlds that are causally and metaphysically inaccessible to us, as it is only fitting to desire those things that are possible ways our world could be. Bykvist argues on this basis that if FA is true, then the solitary egret world is not a good world. But hedonism entails that it is a good world, and thus FA is inconsistent with hedonism. According to Bykvist, a correct analysis of value or goodness must not rule out hedonism as a possible axiology. Therefore, FA is false.

Bykvist’s innovative arguments may well be correct. It is difficult to evaluate them succinctly, because they rely on at least three controversial claims. One is that an analysis of good or value must be sufficiently neutral so as to admit all intuitively plausible axiologies; the failure of FA is the failure to accommodate one such axiology, hedonism. A second is that desire is the central, and perhaps only, relevant attitude for FA. And finally Bykvist assumes a dispositional account of desire with demanding fittingness conditions.\textsuperscript{36}

An additional challenge to evaluating Bykvist’s view, and this is no criticism but rather just an

\textsuperscript{36} For an in-depth discussion of possible problems with Bykvist’s argument, see Orsi (2013) and Bykvist’s reply (2015)
observation, is that his view is not in a strict sense an argument that relies on FA under-generating for value. After all, it may be that some other axiology apart from hedonism is true and that that axiology does not entail that a solitary goods world is a good or valuable world. It is, however, the first argument of which I am aware that suggests that FA could under-generate for even a single plausible axiology and is a substantial innovation in the literature.

It is possible to (try to) produce an under-generation argument against FA that is more neutral in the sense that it does not depend in any interesting way on claims about the ambitions of the value of analysis, but rather shows that on any reasonable axiology, FA under-generates. This is the strategy I pursued in ‘Fittingness, value and trans-world attitudes’.³⁷ This argument also requires various other controversial assumptions, yet it contains the seed of a less controversial strategy for arguing not only against FA, but against at least some other bi-conditional relations between different non-descriptive properties. It is from these seeds that the perhaps the less controversial arguments of this chapter are grown.

³⁷ Reisner (2015a)
material bi-conditionals; if the relevant material 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 \( x \).

\[
\text{F2. RFAV: } x \text{ is valuable } \iff \text{ there is a reason to favour } x
\]

Our target is to develop a schema for creating examples in which \( x \) is valuable, but there is no reason to favour \( x \). We may start by considering the structure of reason relations:

3. The simple reason relation: Fact \( f \) is a reason for agent \( A \) to \( \psi \) to degree \( d \)

In the simple reasons relation, \( [f] \text{act} f \) should be interpreted liberally as including conjunctions of facts or sets of facts.\(^{39}\) The schematic variable \( \psi \) 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.

3.1 The under-generation argument for reasons and value

With the basic 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-

\(^{38}\) Many contemporary writers omit the final place in this relation; as Fogal & Risberg (MS) note, this is a mistake. John Skorupski (2002, 2010) was careful to avoid this mistake in his pioneering work on the metaphysics of reasons. Errol Lord (2018) and other contemporary reasons-first advocates make this mistake.

\(^{39}\) In Skorupski’s (2002) explication of the reason relation, \( f \) stands for a set of facts.
generates.

The simplest structure for such examples relies on descriptive, or if one prefers, non-normative entanglement.⁴⁰ One needs to generate examples in which favouring $x$ causes $x$ not to be valuable.⁴¹ I shall focus on good as a paradigm type of value. Here is a generic counter-example:

4. The generic counterexample: $x$ is good at $t$, 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, however, instead the effect of her doing so will be that she causes pain and suffering around the world. We may treat the effect of her waving her arm in both circumstances as necessary due to her nature.

An example of this form entangles favouring $x$ (descriptive) with $x$’s goodness (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. Favouring $x$ effaces the reasons for favouring $x$ and thus defeats even the weakest guidingness constraints on reasons.⁴²

---

⁴⁰ 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 Włodek Rabinowicz.

⁴¹ Strict covariance is also sufficient.

⁴² See Risberg (2020) and Rosenqvist (2020) for further discussion on guidingness.
Now we are in a position to see why value and reasons have unalike variance conditions in the arm-waving example. So long as nobody favours the powerful entity’s waving her left arm, it is good that she waves her left arm. If somebody favours her waving her left arm, then it is bad 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.

It is important to note that this example can be given without the modally significant claim that it is metaphysically necessary that the entity’s powers work in the way that they do. However, as I shall discuss in §3.2, this additional condition renders one possible objection to the counter-example impotent.

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 experience $x$.

We should understand ‘experience $x$’ 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 climbing its face destroys those natural features that make it beautiful, rendering its beauty impossible to experience.

In this case, presumably the features that make the rock formation beautiful do so whether or
not they can be observed.⁴³ Thus so long as one does not climb the formation, it remains beautiful. But if one is climbing or has climbed the formation, then the formation is not beautiful, due to the destructive effects of climbing it. One has no reason to experience the formation, because doing so effaces the physical features of the formation that provide reasons to experience it; one has no reason (aesthetic) reason to climb the formation once one is climbing it. Here again, we see that one has no reason 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.

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

3.2 Objections to the counter-example schema

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

The first objection is an anti-actualist objection. Consider a well known class of counter-examples to deontic detachment. Deontic detachment is an inference rule that says if you ought to \( x \) and you ought (if you \( x \) then \( y \)), then you ought to \( y \). The counter-example is this. You ought to change lanes and pass the driver in front of you, because the driver in front of you is driving slowly and unsafely. And you ought (if you change lanes and pass the driver in front of you, then you accelerate), because the only safe way to change lanes and pass requires you to accelerate. However, you in fact will not change lanes and pass the driver in front of you. According to deontic detachment, you nonetheless ought to accelerate, but of course doing so will cause you to crash

---

⁴³ 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.
into the car in front of you, since you will not change lanes. Deontic detachment thus entails that one is doing something that one ought to do by accelerating into the car in front of one’s own.

There are sceptics of this sort of argument.⁴⁴ They defend for various reasons possibilism about deontic inferences. Because it is possible for you to change lanes and pass the car, it remains true, according to them, that you ought to accelerate. You cannot cancel your obligations simply by making choosing to behave in ways that would make filling the obligation counterproductive.

While the debate between actualists and possibilists is interesting, the arm-waving example resists the possibilist strategy. In the example, it is part of the arm-waving entity’s nature – it is metaphysically necessary that – her waving her left arm relieves pain and suffering when (eternally) nobody favours it and causes pain and suffering if anyone at any time favours it. Possibilism is plainly not an available response to that sort of entanglement case.⁴⁵ It may be a good response to cases where the entanglement is contingent rather than necessary, although I remain sceptical about possibilism more generally. Of course, one could attempt to reinsert necessity in less metaphysically exotic ways by generating examples in which the entanglement is causally necessary.

A second objection is 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 been suggested to me with respect to fittingness. There are a number of technical issues that arise in with respect to this proposal, many of which I have discussed in depth in an earlier paper.⁴⁶ However, I am now convinced that there is a (somewhat) more straightforward way to reply to this objection, at least with respect to reasons.

⁴⁴ Most importantly Benjamin Kiesewetter (2018).
⁴⁵ The inspiration for this example comes from a conversation with Ralf Bader about potential problems with the principle of necessary detachment.
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.

Although the arm-waving case is stated in general terms, if 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 left arm (when nobody favours her doing so generally) is also good. A successful of analysis of good, 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 nobody (eternally) 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 of the entity’s waving her left arm. This is because reasons are indexed to individuals. It is doubtful that individuals in other possible worlds can favour an entity in the actual world’s doing so, 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; or at least there are good arguments to suggest that it is metaphysically impossible to have singular thoughts about individuals or occurrences in other worlds.\(^{47}\) While it is an open question as to whether \textit{reason} implies \textit{can} for some senses of \textit{can}, one seems to have lost all grip on the notion of a normative reason if \textit{reason} does not at least imply \textit{can} \textit{metaphysically}.

\(^{47}\) \textit{Ibid.} and see Soames (2002).
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 world⁴⁸ 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 world is good in her world.

Of course talk about what possible rather than actual individuals have reason to favour in the actual world is strange in numerous ways, not least of which because it is barely intelligible without accepting modal realism. The very claim that \( x \) is good if a possible person favours it sounds false; it is difficult to know how to make sense of possible people in a way that would lend even some plausibility to the proposal without accepting modal realism. Strangeness aside, PRFAV is extensionally inadequate, which is enough to reject it without complaining about the metaphysics.

The final objection concerns the ‘eternity’ condition in the counter-example 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 the assumption that the future is open and thus not knowable. I do not have very much to say about this objection, because it clearly hinges on the difficult question of whether the future is knowable and whether perhaps on whether time-travel is at least metaphysically possible. I am content to let this counter-example be held hostage to such concerns.

---

⁴⁸ The \textit{actual} operator indexes to this world, where as ‘another possible world’ should be taken to contain a different indexical operator \( W \), that functions like the \textit{actual} operator but localises to the world in which it is being used. I discuss how this operator works in Reisner (2015).
for the time being. But I would like to stress that I take it seriously.

4. THE ARGUMENT GENERALISED TO FITTINGNESS

If the argument in §3 is correct, then the remainder of the argument against at least simple versions of all three-first views is straightforward:

1) Value and reasons have unalike variance conditions
2) If two categories of properties have unalike variance conditions, then neither can be analysed in terms of the other.
3) Sub-conclusion: neither reasons nor value can be analysed in terms of the other.
4) Fittingness either shares variance conditions with one, but not both, of reasons or value, or it shares variance conditions with neither.
5) Sub-conclusion: fittingness cannot be used to analyse both value and reasons.
6) Conclusion: None of fittingness, reasons, or value may be used to analyse the other two non-descriptive properties.

This argument is valid, and it applies even to more sophisticated -first accounts, like Christopher Howard’s,⁴⁹ which is cleverly constructed so as to avoid cases where fittingness under-generates for reasons.

Howard’s concern is that if we accept that there are state-given reasons for propositional attitudes, standard fittingness first theories under-generate. My presentation of Howard’s view is not entirely faithful to the original, but the changes affect small details in the presentation and not the central extensional adequacy concerns. His account is built on two main claims:

⁴⁹ Howard (2019)
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 desire $x$ if and only if: 1) it is fitting to desire $x$, or 2) it is fitting to desire that one desire $x$.

VAF is just FA. RAF, read with the first conjunct alone, says that there is a reason to desire $x$ if and only if it is fitting to desire something. That would appear to rule out state-given reasons to desire, like those given in Crisp’s bowl-of-mud example. Intuitively, it is good in that example that you desire the bowl of bud. That entails, according to VAF, that it is fitting to desire that you desire the bowl of mud. Howard stipulates that when a second order desire it 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 if and only if one has a reason to desire or favour it. That is because 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 $x$.

The arguments in §3 show that reasons have unalike variance conditions from value. This raises a new question. Which non-descriptive property has unalike variance conditions from those of fittingness? The answer, I shall argue, is at least value. To do so, I shall introduce a new version of WKV for fittingness. I shall take up the question of whether reasons and fittingness ever have unalike variance conditions in §5.

4.1 *The over-generation argument 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 argument presented in §3 against RFAV. One need only swap in ‘fittiness’ for reasons and fix the grammar accordingly to see how such an argument would look.

However, there is a complication. The argument in §3 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) φ whilst there is (still) a reason for A to ψ.

As I noted in §3, it is difficult to doubt this condition. 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 to favour A’s ψ-ing only if A can (metaphysically) ψ whilst it is (still) fitting for A to ψ.

Intuitions about this principle may be less clear than they are for RCR. However, I do not think that RCF is really doubtable. Consider the following value claim and its fitting attitude counterpart:

C1V. It is good that A ψs only if A does not ψ.

C1F. It is fitting to desire that A ψs only if A does not ψ.

C1f follows from C1v, since the conditions on fittingness must be the same as the conditions on value. C1v is surely false on at least one reading. It is an instance of an analytically false general
schema:

C₁G: There is some instance of A’s ψ-ing such that there is no instance of A’s ψ-ing and A’s ψ-ing is F

This reading concerns the properties of a particular event, rather than an event type. Particular events have properties. In order to have those properties, they must exist. C₁V and C₁F are therefore false when read as being about particular events.

Of course, that is not the only reading. However, the reading does not improve if we treat A ψs as a comparison class. It still follows that A’s ψ-ing is never good, and thus the unconditional claim, that A ψs is good, is always false.

We can leverage this insight into a defence of RCF:

C₂F. It is fitting that S favour A’s ψ-ing only if S does not favour A’s ψ-ing.

There is no instance of S’s favouring A’s ψ-ing that has the property of being fitting, if this conditional is true. We can strengthen C₂F to share the same modal robustness as we see in the arm-raising example:

C₂F*. 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.

According to C₂F*, not only are there no actual instances of S’s fittingly favouring that A ψs, but there are also no possible instances of it either. There are no actual or counterfactual instances of
S’s fittingly favouring that A ψs.⁵⁰

Someone who wishes to deny RCF must offer an interpretation of ‘fitting to favour’ that is consistent with there’s being no possible instances of favouring, actually or counterfactually, that have the property of being fitting.

Insofar 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,⁵¹ 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-airs 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. The first is the example of attitudes that are fitting on comparative grounds:

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 Rocky, Rocky Balboa watches a fight on t.v. from 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, noting 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. The retort and the admiration would be out of place if Creed’s efforts were merely typical for

---

⁵⁰ Note that a similar situation arises in Bykvist’s solitary goods world; there are no individuals who can favour the egret’s being happy, nor could there be, assuming that the world’s being populated by a single egret is essential to its being the world that it is.

⁵¹ See §3 and Reisner (2015).
Rocky, the bartender, and the other 'schlubs' from the neighbourhood.

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. $S$ comes to learn about $A$, who exists in a different possible world. Although $A$ is in fact a pretty awful person by the standard’s 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 feats of daring-do in other worlds. In her own 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 her doing so would cause. But let us suppose that $A$’s climbing the mountain in her own world will allow her to install the relay that will bring Skynet online. Thus her climbing in the mountain is bad in her own world. It is fitting for $S$ to favour that $A$ climb the mountain, but it not good in $A$’s world that she do so, violating the central bi-conditional of FA.

I therefore conclude that value cannot be reduced to fittingness, and I have argued that value cannot be reduced to reasons. This entails that value is not subject to analysis or bi-conditional equivalence in the manner required for fittingness-first and reasons-first theories.

5. 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 return to Danielsson & Olson’s 2007 paper.

52 Henry Hill expresses a similar sentiment, although in his case about being a schnook, in Goodfellas.
When Danielsson & Olson set out to solve the wrong kind of reason problem, they did so by importing an additional non-descriptive notion, *correctness*. Correctness is fittingness. Recall that their strategy was to divide normative (holding) reasons into two kinds: those that arise from correctness and those that do not. The former are suitable for FA, and the latter are not. The reasons that are not suitable for use in FA cannot, according to Danielsson & Olson, be accounted for by correctness. Or to be more precise, they do not claim that those reasons are accounted by correctness.

If we look back to Howard’s *reasons as fittingness* condition, he offers a way of accounting for the non-correctness reasons in terms of fittingness. 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-correctness* reasons. Let us suppose that it is.

A proposed advantage of Howard’s view is that it offers conceptual gain. 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.

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 value. Reasons’ sensitivity to value is explained by the underlying fittingness 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 gain 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.

To this end, I am much more strongly inclined to think that a view like that offered by Conor McHugh and Jonathan Way\(^3\) is well motivated, albeit still 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 correctness reasons. And this fact is meant to be explained by the primacy of fittingness.

If one assumes, as I do, that there are non-correctness reasons for pro-attitudes, McHugh & Way’s account is not extensionally adequate. Howard’s account provides an extensionally adequate bi-conditional linking fittingness and reasons. But there are two grounds for doubting that this suggests that fittingness is prior to reasons.

The first of these grounds is that fittingness does not do the work of reasons. The idea that underlies reasons, oughts, and other normative – in the narrow sense used in this chapter – properties is that they are guiding in some loose sense. This seems to put a fittingness-before-reasons view onto the horns of a dilemma. It fittingness is not a normative property, then there is more to something’s being a reason than its being fitting; a new feature, guidingness, is added. On the hand, if fittingness is guiding, 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.

If the arguments in the rest of this chapter are correct, and fittingness is not prior to value,

---

\(^3\) McHugh & Way (2016)
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.

This is clearly not a conclusive argument against the claim that fittingness is prior to reasons. However, normative notions are central to much of our ethical and even epistemological theorising, and if we are not willing to lose the guidingness that characteristic 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 here.

6. Conclusion

In this chapter I have argued that value is not analysable in terms of reasons or fittingness, due to the extensional inadequacy of such analyses. This also shows that -first views that have the ambition to reduce two of fittingness, reasons, and value to the remaining third property or concept are false. I have not taken up the interesting question of the aims or ambitions of -first projects. If the argument 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 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 §5 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.
Works Cited


Olson, Jonas (2009a), ‘The Wrong Kind of Solution to the Wrong Kind of Reason Problem’,


