ANDREW REISNER

25 JANUARY 2023

<div align="center">

V. AGAINST THE 'FIRST' VIEWS.

</div>

[Chapter 5 of *A New Theory of Reasons for Belief: Pragmatic Foundations and Pluralistic Reasons* under contract with OUP.] This draft is full of typographical errors. It is currently undergoing significant revisions, but it is roughly along the lines of what should end up in the book. Please check with me before citing.]

0. INTRODUCTION

Future historians of philosophy looking at the present times may note the proliferation of work directed at finding which normative or evaluative concept or property is first.[1] In a broad sense, one might see such projects as part of an effort to explain why there are at least apparent links between properties such as *being a reason to favour, being fitting to favour, and being valuable*[2] or as part of projects with particular metaphysical aims such as maximising ontological parsimony.[3] Individual positions such as *reasons first*, *fittingness first*, and *value first*[4] have all individually been

---

[1] Undoubtedly part of the explanation for this is philosophical. Perhaps part of the explanation is also sociological, namely 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. I do not mean to suggest that this explanation applies to any particular philosopher discussed in this chapter.

[2] This explanation was suggested to me by Daniel Star.

[3] While it is good practice not to invent new categories of properties without proper motivation, the commonly used non-descriptive categories discussed in this chapter map onto intuitive concepts that, at least at first blush, seem to be distinct enough as not to beg for reductive analysis. Whatever pressure there is to aim for maximum ontological parsimony, it must be balanced against the naturalness and general plausibility of the reductive analyses employed to achieve it. As I am not at all persuaded that *-first* views have substantial explanatory value, it is my view that their attraction is superficial.

[4] Several authors defend the fittingness-first approach. Some important examples include: Chappell (2012), McHugh & Way (2016 & forthcoming), and Howard (2019). Reasons-first views perhaps receive the most prominent treatment in the literature. For book-length treatments of different versions of the view, see Lord (2018), Schroeder (forthcoming), and Skorupski (2010). For further discussion of the motivations behind reasons-first views, see also Schroeder (2021). A version of value-first is defended in Maguire (2016) and perhaps has antecedents in the work defending scalar

criticised, often by proponents of one of the other types of what I shall refer to as '-first' views. I shall call the properties and concepts in the broad category that includes *being a reason*, *being fitting*, and *being valuable* 'non-descriptive'. This chapter has a broader goal and develops an argument aimed at showing that at least straightforward versions of each of these -first views about non-descriptive properties is false.[5]

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 certain *-first* views – reasons, fittingness, and value – have 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 a change in descriptive properties will lead to a change in a non-descriptive property in one category (e.g. in the evaluative category from good to either bad or neutral, or in the fittingness category from fitting to fitting not to) in circumstances in which non-descriptive properties in other categories do not change.[6]

Showing that different non-descriptive 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-extensive, which is the minimum condition for an analysis, reduction, or an identity claim.[7] For the most part, the examples of unalike variance

---

utilitarianism, for example Norcross (2006).

[5] There are of course other possible *-first* views. For an example of another way that projects in this vicinity might be approached, see Tännsjö (2008) for an interesting discussion of *ought* as the central moral notion.

[6] I have revised the description of the strategy in light of comments from Jonas Olson. It is important to highlight that this strategy is not new. An argument explicitly described much this way can be found in Heathwood (2008), and related arguments appear in Bykvist (2009) and Reisner (2015).

[7] 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's being any correct -first view. Daniel Wodak (forthcoming) argues from broader metaphysical considerations against the general -first project in the non-descriptive domain. I am setting aside complications to the claim about necessary bi-conditionals being required for analyses raised by cases of multiple-realisability. I thank Krister Bykvist for reminding me about this point.

conditions are instances of under-generation, but some examples are instances of over-generation.[8]

I have argued elsewhere[9] that fittingness and value vary in unalike ways under various pairs of circumstances. I concluded from those arguments 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 first 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 set of arguments that fittingness and value have unalike variance conditions. If these arguments are right, then value is one of at least two non-descriptive primitives[10] on the perhaps false assumption that value, reasons, and fittingness are the three best candidates for being the non-descriptive primitives.

This much leaves open the possibility that at least one of the non-descriptive properties may be reducible to one of the others or to a combination of the other two. I tentatively suggest that reasons cannot be reduced to or analysed in terms of fittingness, leaving fittingness as the sole possible candidate for analysis or reduction. However, I do not explore the matter further.

There is much more to say about the topics discussed in this chapter than is said. The rôle of this chapter within the broader context of this book is narrow, but important. The central theory that I advance concerns the relation between a particular family of accounts of wellbeing and their relation to the considerations that determine what one ought to believe. This relation, or so I posit, is one in which different components of wellbeing contribute to determining the total value of an individual's wellbeing, and that ultimately what one ought to believe is determined by the value of holding different beliefs. This picture, as I understand it, proves not to fit very naturally with either

---

[8] Sarah Stroud rightly urged me to point out that there is nothing more to unalike variance conditions for two non-descriptive properties than that they are not necessarily co-extensive.

[9] See Reisner (2015).

[10] It is important to distinguish this part of the project from the work of authors like Maguire (2016), whose aims are different and in my view rather less plausible.

reasons-first views or fitting-attitude accounts of value. Philosophically, I have no objection to the possibility of having a richer basic non-descriptive[11] ontology that contains more than one category of non-descriptive property, e.g. normative (reasons/oughts) properties and value properties. This chapter provides an argument that at least evaluative and normative properties are basic and also distinct from each other. If the arguments herein concerning extensional adequacy are correct, then that is sufficient to show that evaluative and normative properties are distinct and basic, if there be any such properties at all and if those properties are the only -first candidates under consideration.[12]

1. An overview

Moral philosophers are accustomed to working with two at least apparently distinct families of broadly non-descriptive properties, namely evaluative properties and deontic properties.

Evaluative properties be thin, such as *good* or *bad*. Alternatively they may be thick, such as *admirable* or *cruel*. Moral deontic properties, including *right*, *wrong*, and *morally permitted* are often understood as being special cases of non-moral deontic properties, namely *ought*, *forbidden*, and *permitted simpliciter*.[13] We may for convenience group deontic properties under the wider umbrella of normative[14] properties, which also includes *normative reasons*, but which does not include either evaluative properties or correctness properties such as *fittingness*.[15]

---

[11] Jonas Olson pointed out to me that 'non-descriptive' can sound as though I am partial to some form of non-cognitivism. This is not my intention. I assume cognitivism here and throughout the book.

[12] My thanks to Hichem Naar for his suggestion that I expand the introduction to explain the chapter's rôle in the book.

[13] See Zimmerman (2015) for more discussion.

[14] 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. I thank Daniel Star for pointing out that my earlier usage of the term was inadequately explained.

[15] '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 manner I

Some substantive moral theories are expressed in terms of relations between evaluative properties and normative 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 definition of *ought*,[16] in general work on consequentialism makes no special claim at all about whether normative properties can be analysed in terms of evaluative ones or *vice-versa*. Nonetheless, ethics scholars of many different stripes are disquieted by the lingering possibility that at bottom there are unanalysable, basic non-descriptive properties. While many scholars perhaps reluctantly accept that there is at least one type of unanalysable non-descriptive property, they maintain a the-fewer-the-better approach.[17] For reasons that are not well-explained by the authors themselves, it seems that evaluative properties have been a particular source of anxiety.[18]

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*.[19] Let us focus on the metaphysics. In the current literature, the property of *being fitting* is typically understood as being either its own basic type of property or as being identical to the property of *being a normative reason*, or at least a certain type of normative reason.[20] The so-called *fitting-attitude analysis of value* (FA) can be stated thus:

---

stipulate.

[16] Or more accurately, *right,* e.g. Moore (1903). I thank Jonas Olson for correcting a mistake in an earlier draft.

[17] Wlodek Rabinowicz calls this 'conceptual gain', see Rabinowicz (2008 & 2012).

[18] This point emerges clearly in Lang (2008).

[19] See Danielsson & Olson (2007) for a detailed discussion of the history of the fitting-attitude analysis of value. Kriegel (2018) provides an in-depth study of Brentano's systematic philosophy, including his work on fitting attitudes and value. Nils Sylvan (2021) has a comprehensive catalogue of versions of the fitting-attitude analysis.

[20] As Krister Bykvist and Jonas Olson both pointed out to me, there are several other historical proposals, but I shall confine the discussion to this one, which has dominated the literature on the subject over the last two decades.

1. What it is for x to be valuable is for it to be fitting to favour x.

which entails (amongst other things) a simple bi-conditional equivalence:

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

In 1a) 'favour' is a generic pro-attitude verb. We may treat this as a stipulated usage of 'favour'. In this instance, the formulations should be read neutrally insofar as 'fitting' may be understood either as a distinct property from that of *being a (certain type of) reason*.

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*.[21] The details of the buck-passing account need not detain us here. More importantly for present purposes, with the subsequent publication in 2004 of Wlodek 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

---

[21] Scanlon (1998). For a helpful discussion of what buck-passing really amounts to, see Olson (2009). Dancy (2000) offers an important contemporaneous discussion of Scanlon's buck-passing account.

amongst others and later as a *sui generis* property by several other authors).[22]

It is important to highlight two of the purported advantages of the account. The first is what Wlodek Rabinowicz calls 'conceptual gain'.[23] 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 early interest in conceptual gain presages the more comprehensive -first projects.

The second purported advantage is that relations amongst reasons or fittingness and value are, if not explained, at least accounted for.[24] Value 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.[25] It suffices, because it helps to explain how the -first views in many cases arise.[26] 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 property to treat as fundamental, or first, one increases the conceptual or metaphysical gain and also extends

---

[22] E.g. Danielsson & Olson (2007). Rabinowicz & Rønnow-Rasmussen also consider this possibility, only to reject it as problematic for purposes of analysis. Their central concern is that one would be replacing one primitive, *value*, with another, *fittingness*.

[23] See fn. 14.

[24] I discuss difficulties with this kind of explanation in Reisner (2009a).

[25] *Ibid*. One may indeed wonder, as I do, whether they really are advantages.

[26] I do not mean to suggest that this is the only way in which -first projects may arise. For example, Nathan Howard pointed out to me that Schroeder's primary motivation for defending the reasons-first programme is because reasons in his view are consistent with the kind of synthetic naturalism that he favours.

the explanatory(-ish) project.[27]

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*,[28] or of apparently descriptive but problematic ones such as modality and probability.[29] The strategy of finding unalike variance conditions is fully generic, and perhaps it may be fruitfully applied to these and other extended 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 properties. The two properties with conditions that are most clearly unalike 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 should like to revisit the now 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 property. I shall argue that FA fails in §4, but for the moment shall rest content to make some salient methodological observations that will provide context for the rest of the chapter. The discussion here only concerns non-descriptive properties, but both the discussion and the arguments it contains may for the most part be freely adapted to apply to the corresponding non-descriptive concepts.

---

[27] Strictly speaking, one could accept that a putatively non-descriptive property was both first and non-fundamental, if for example that property were to be reducible to one or more descriptive properties.

[28] Lord (2018)

[29] Skorupski (2010).

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 a terminological variant of 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 problem* (WKR). Lost to some extent today is the original usage of this term, deriving from Rabinowicz and Rønnow-Rasumussen's 2004 discussion[30] of WKR.[31] 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 many of the other contexts in which it is now used,[32] but the possible defects of the post 'Strike' literature lie outwith the present discussion. The original invocation of the wrong kind of reason arises for an understanding of 'fitting to favour' as 'there's 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[33] if and

---

[30] Rabinowicz & Rønnow-Rasmussen (2004)

[31] Pamela Hieronymi's (2005) discussion of right and wrong reasons for use as premises in reasoning seems to have initiated the perhaps regrettable change in practice.

[32] For examples of the current usage, see Gertken & Kiesewetter (2017) and Sylvan & Lord (2019). In some 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 in some generic sense. Of course, there may be reasons of the right and wrong kind for other analyses apart from FA, as Nathan Howard and Krister Bykvist both pointed out to me. Howard suggests the example of something's being frightening if and only if there is a reason to fear it and Bykvist of someone's being blameworthy if and only if there is a reason to blame her. However, the usage of 'right reasons' and 'wrong reasons' seems to be wider than that; for example Sylvan & Lord do not identify (even) right reasons with normative reasons. I thank Daniel Star for his comments about this issue.

[33] This use of 'valuable' is intended to be consistent with a version of RFAV that seeks to identify something's being valuable in a respect, even if it is not valuable overall.

only if there is a reason to favour x.[34]

Drawing on an example of Roger Crisp's in which an evil demon conjures us to desire[35] a bowl of mud despite the undesirability of the bowl of mud itself,[36] 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.[37] 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 Piller[38] 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 attitude and its contents)[39] 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

---

[34] In order to achieve conceptual gain, RFAV must be read in line with its name as an analysis, and thus at least 'with determination going from right to left' should be added. For present purposes it is enough to work with the simple bi-conditional.

[35] I use 'desire' as a specific favouring attitude for purposes of the example.

[36] Crisp (2000).

[37] 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.

[38] Piller (2006) uses a different nomenclature, distinguishing instead between 'content-given reasons' and 'attitude-given reasons'. The standard nomenclature of 'object-given' and 'state-given' originates to the best of my knowledge in Parfit (2001).

[39] See Danielsson & Olson (2007) and Reisner (2009a, 2014 & 2018) for more discussion.

another, developing the *pairing recipe*,[40] which appears to resist the state-given reasons as wrong reasons and object-given reasons as right reasons solution to WKR. There has been much subsequent literature on WKR and possible solutions[41] 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',[42] in which they propose to identify the right kind of reasons with what they call 'content reasons'. A content reason is not itself always normative reason. Those are called 'holding reasons' by Danielsson & Olson, but content reasons, according to Danielsson & Olson, (almost) always give rise to holding reasons, although not all holding reasons have directly corresponding content reasons. Content reasons are facts that make an attitude correct in virtue the relation between the attitude type and its contents. For example, that x has certain properties – properties that make x desirable – is a content reason to desire x. That I would get a prize for desiring x is a holding reason for desiring x, but not a content reason, as that does not make x itself desirable.[43]

In this picture, holding reasons that lack corresponding content reasons are reasons of the wrong kind for RFAV, whilst those that have corresponding content 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. Whether its being correct (having a content

---

[40] This is the name that Gerald Lang (2008) uses for it. Rabinowicz & Rønnow-Rasmussen (2004) develop a generic recipe to convert state-given reasons into object-given reasons, thus rendering the distinction unhelpful for solving WKR: 'If a pro-attitude towards an object a would have a property P, then, ipso facto, a has (or would have, if it existed) the property P' of *being such that a pro-attitude towards it would have the property P*. Consequently, to the attitude-given reason, provided by P, corresponds the object-given reason, which is provided by P'. In exactly the same way, of course, for any property P of the object of the attitude there is a corresponding property P' *of the attitude itself: the property of being such that its object has (or would have) property P*. Thus, to each object-given reason corresponds an attitude-given reason, and vice versa.' (page 406).

[41] See Lang (2008), Olson (2009), and Samuelsson (2013) for further discussion.

[42] Danielsson & Olson (2007).

[43] I thank Jonas Olson for some further guidance on how to interpret Danielsson & Olson (2007). Their account contains further nuances. I refer the reader to the original paper to explore the matter further.

reason) to favour *x* is the same thing as there's being a reason (in the standard sense of the term) to favour *x* depends on whether one is convinced by their Ewing-inspired analysis of holding reasons given in terms of content reasons.[44] Since it remains an open question in this chapter as to whether such an analysis can be successful, I should instead like to consider the importance of the content/ holding reason distinction on its own. If we take Danielsson & Olson for the sake of argument to have solved WKR in this way, then we still see conceptual (or rather ontological) gain in the basic non-descriptive ontology with both only fittingness and normative reasons as essential parts.

We now have a picture according to which one no longer needs *per se* to employ normative reasons directly in the analysis of value. Rather, it is enough that we can (let us suppose) create an extensionally adequate bi-conditional linking fittingness (now replacing 'content reasons') and value without making 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.[45] 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*.

At least four authors have been involved in developing what one might call the 'wrong kind of

[44] As Nils Sylvan and Jonas Olson both pointed out to me.
[45] Crisp (2005) and Reisner (2009a)

value problem' (WKV). The original idea, which seems not to have taken hold despite its promise, is due to Jonathan Dancy.[46] His work was followed by Christopher Heathwood,[47] in a paper in which he develops an argument from unalike variance conditions over time.[48]

The next and most influential contribution to the WKV literature is due to Krister Byvkvist[49] 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 future possessed of a mental life. We should assume that the egret lacks the requisite psychology for having fitting attitudes. According to hedonism, for example, 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. This kind of case can be made to work with a wide range of axiologies.

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 in fact to 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. The reasons for this vary according to the particular type of attitude that favouring may be taken to stand for.[50] Like Heathwood, Bykvist aims to deliver a general argument against FA by looking for cases of under-generation.

---

[46] Dancy (2000).

[47] Heathwood (2008). I thank Jens Johansson for directing me to this important paper. I regret that I was not aware of it when writing on this topic in the past.

[48] See Johansson (2009) for thoughtful criticisms of Heathwood's view.

[49] See Bykvist (2009).

[50] For an in-depth discussion of possible problems with Bykvist's argument, see Orsi (2013) and Bykvist's reply (2015). I thank Bykvist for correcting several misunderstandings of his view in an earlier draft.

This is also the strategy I pursued in 'Fittingness, value and trans-world attitudes'.[51] In contrast to Bykvist's central example, which concerns attitudes we might have from our world towards individuals in another, putatively good world, I focused on instances of value in the actual world (from our perspective). The arguments in that paper rely mainly on the effects that having certain pro-attitudes can have on the value of their objects. Parts of the argument rely on complex questions concerning modal metaphysics and thought. The arguments in this chapter are in a broad sense a development on those arguments. However, I have as much as possible avoided relying on any controversial metaphysical claims and have tried find simpler and more direct ways of arguing in a related manner. I have not in all instances succeeded.

3. Unlike variance conditions for reasons and value

Reasons and value have unlike variance conditions, or so I shall argue. And if they have unlike variance conditions, then that is enough to show that no straightforward -first theory will be correct.

All -first views, or at least any -first view with the ambition of analysis or 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:

---

[51] Reisner (2015)

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 $\longleftrightarrow$ there is a reason to favour x[52]

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[53]

In the simple reasons relation, '[f]act f should be interpreted liberally so as to include conjunctions of facts or sets of facts.[54] 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.

3.1 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

---

[52] As Velislava Mitova has pointed out to me, the reason relation on the right side of this conditional is only a two-place relation. I have left it this way for ease of presentation, but see fn. 50 for some further remarks.

[53] Many 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. Errol Lord (2018) and other contemporary reasons-first advocates make this mistake.

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

entanglement.[55] One needs to generate examples in which favouring x causes x not to be valuable.[56] I shall focus for now on *good* as a paradigm type of value. Here is a generic counter-example:

> 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 due to her nature.[57]

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. Favouring x effaces the reasons for favouring x and thus defeats even the weakest guidingness constraints on reasons.[58]

---

[55] 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.

[56] Strict covariance is also sufficient.

[57] In the past (2009a 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.

[58] 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

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.[59] I shall consider three possibilities: intrinsic final value, extrinsic final value, and instrumental value. 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 had 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.[60] 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 this the arm-waving example concerns extrinsic final value. It should be much less controversial to say that the arm-waving example is a

---

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.

[59] 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.

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

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.

As second example is required to create clear difficulties for an analysis of intrinsic final value.[61] Let us suppose, as many philosophers have, that it is intrinsically finally valuable to love another person unconditionally. I am assuming that 'intrinsic' should be understood broadly enough to include agents with other-involving mental states.

This example also involves a demiurge who decides this time that if anyone ever favours a particular instance of Xenophon's unconditionally loving a 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, because the existence of the pair {Xenophon loves x unconditionally at t, anybody ever favours that Xenophon loves x unconditionally at t} is impossible. No instance of Xenophon's loving another 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 us to experience.[62]

In this case, presumably the features that make the rock formation beautiful do so whether or not they can be experienced.[63] 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. 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.

It bears noting at this point that although RFAV is formulated as a simple 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 that self-effacing (putative) reasons to favour

---

[62] 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. I do not have a convincing response to either worry.

[63] 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.

are not actual reasons to favour is correct. I believe it is, but I should 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 car in front of you, because its driver is driving slowly and unsafely. And you ought (if you change lanes and pass the car in front of you, then you accelerate), because the only safe way to change lanes and pass the car requires you to accelerate. However, you in fact will not change lanes and pass the car in front of you. According to deontic detachment, you nonetheless ought to accelerate, but of course doing so will cause you to crash 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. This seems to be clearly wrong.

There are sceptics of this sort of argument.[64] They defend for various reasons *possibilism* about at least some 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 choosing to behave in ways that would make filling the obligation counterproductive.

While the debate between actualists and possibilists is interesting, the necessary version of 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.[65] It

---

[64] Most importantly Benjamin Kiesewetter (2018).
[65] The original inspiration for this discussion comes from a conversation with Ralf Bader about potential problems with the principle of necessary detachment.

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 rather than a question of some being's nature.

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 with respect to this proposal, many of which I have discussed in depth in an earlier paper.[66] However, I am now convinced that there is a (somewhat) more straightforward way to reply to this objection, 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.[67]

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 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

---

[66] Reisner (2015).

[67] 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.

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,[68] which is what would be required to get a particular individual (situated in a particular world) in mind. While it is an open question as to whether *reason* implies *can* for some senses of *can*, one seems to have lost all grip on the notion of a normative reason if *reason* does not at least imply *can metaphysically*.

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[69] performs a cruel act that causes only pain.[70] 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.

---

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

[69] The *actual* operator indexes to this world, where as 'another possible world' should be taken to contain 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).

[70] Presumably for this discussion to make any sense at all, one would have to accept some form of modal realism.

Of course talk about what merely possible rather than actual individuals have reason to favour in the actual (from out point of view) world is strange in numerous ways, not least of which is because it is at best barely intelligible without accepting modal realism. The very claim that x is good if a merely 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 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.

4. The argument extended to fittingness

If the argument in §3 is correct, then reasons-first is ruled out. This still leaves the possibility of 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.[71] The problem is that fittingness seems to under-generate for reasons, at least if one accepts that there are state-given reasons for propositional attitudes. Howard's account is

---

[71] Howard (2019)

cleverly constructed so as to avoid cases where fittingness under-generates for 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.[72] 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, e.g. 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 mud. 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 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 §3. They show that reasons under-generate for 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. It is easier to see this when written down in a step-by-step way:

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 §3)

5) Conclusion: Either 1 or 2 is false   (from 3, 4)[73]

This raises a problem for Howard's view, namely that either VAF is false or that RAF is false, and thus 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 of WKV for fittingness. I shall take up the question of whether reasons and fittingness have unalike variance conditions in §5.

4.1 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 §3 against RFAV. One need only swap in 'fittigness' 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

---

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

'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 in §3, it is difficult to doubt this condition, which may be understood as an extremely weak guidingness constraint.[74] 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':

> 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.

---

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

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.[75]

F1-F4 would be pretty odd things to say. Presumably, they are odd to say because they each imply a conditional claim of the form: If you will feel x towards A, it will not be fitting/appropriate/correct/meet to feel 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 to then 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. 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.[76]

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 both C1F and C1Fp are false. Indeed, one need not appeal directly to RCF to

---

[75] Thanks to Jimmy Goodrich for suggesting a valuable revision to these examples.

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

see that on all reasonable readings, they (probably) are false. However, the falsehood of C₁F and C₁Fp depend on RCF's being true. While it is not necessary to invoke RCF in arguing against C₁F and C₁Fp, it is important to note that an argument against C₁F or C₁Fp is in essence a direct argument against RCF.

One unclarity about C₁F and C₁Fp is how to understand 'favouring A's ψ-ing'. All readings are problematic. One way to read it is with a universal quantifier: all favourings of A's ψ-ing are fitting for S. But C₁F and C₁Fp entail that no favouring of A's ψ-ing are fitting; on this reading they are simply false.

Another possible reading 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 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 features 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.

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 inherit their properties from their types. Given that there is no difficulty with there's being 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

should not be regarded as rendering 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 inherit heritable properties and relations from the type itself, where the failure to inherit is not due to category problems.[77] To the best of my knowledge, there has been very little work done on spelling out the conditions under which tokens inherit properties or rôles in relations from their types, and thus I make the foregoing comments with all due hesitance.

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 there's being 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,[78] 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.

---

[77] I thank Louis deRosset and Matti Eklund for very helpful correspondence on the question of the inheritance of properties and relations between types and tokens.

[78] See §3 and Reisner (2015).

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:

> 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 the movie, *Rocky*, Rocky Balboa is watching a fight on t.v. 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.[79] The retort and the admiration would be out of place if Creed's efforts were merely typical of those made by Rocky, the bartender, and the other 'bums[80] 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 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

---

[79] A point made clear in the temporally distant sequel, *Creed*.

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

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 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 allow 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 not good in *A*'s world that she do so, violating the central bi-conditional of FA.[81]

I therefore 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 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.

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. Recall that 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.

---

[81] 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 myself.

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, and thus they are not themselves normative reasons, 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 the non-correctness reasons in terms of fittingness. Because I have already introduced Howard's account in some detail, I shall mainly focus on it in the remainder of this section.

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. 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 value. Reasons' sensitivity to value is explained by the underlying 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 alone.

To this end, I am much more strongly inclined to think that a view like that offered by Conor McHugh and Jonathan Way[82] 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.[83] 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 idea 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. 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 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

---

[82] McHugh & Way (2016 & forthcoming).

[83] 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.

and its contents, which is not indicative of fittingness's being prior 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.

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 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 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 category are false. I have not taken up the interesting question of the 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 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.