Fitting Attitude Theories of Value

At the most general level, fitting-attitude theories of value (FA theories) seek to explain various kinds of evaluative facts by appeal to normative facts about certain types of human responses. FA theories can differ according to (i) the kinds of evaluative facts they aspire to explain, (ii) the kinds of normative facts about human responses in terms of which they explain the evaluative facts, and (iii) the types of responses to which they appeal in their explanations. For example, an FA theorist might propose to explain facts about who's *admirable* in terms of facts about the *fittingness* of *admiration*: they might hold that whenever someone is admirable, that's fully because that person is fitting (for just anyone) to admire. And they might offer parallel explanations, in terms of the fittingness of various other types of response, for a range of other evaluative facts, such as facts about who's lovable, blameworthy, despicable, delightful, etc.

Explanations like these are the bread and butter of FA theories. It seems platitudinous that someone is admirable just in case they're fitting to admire, lovable just in case they're fitting to love, blameworthy just in case they're fitting to blame, and so on. What makes these claims of equivalence and their kin so compelling is that the evaluative properties they reference each seem clearly connected to a human response of a certain kind where the relevant connection seems normative rather than merely descriptive. Being lovable, for instance, isn't merely a matter of being someone whom one *does* or *could* love; instead, it's a matter of being someone who's *fitting* to love or who *merits* this attitude. And the same goes for many other specific evaluative properties, including but not limited to those mentioned above (Brandt 1946, Schroeder 2010, Way 2012, Howard 2018). FA theories offer a natural account of these equivalences. For example, the FA theorist can say that what explains the equivalence of *being lovable* with *being fitting to love* is that facts involving the former property are always fully grounded in facts involving the latter.

Many FA theorists aspire to explain not only various specific evaluative properties, such as being lovable or delightful, but also generic value, or the property of being good or valuable *simpliciter*. Assuming that being valuable *simpliciter* is identical to being desirable, an FA theory is straightforward: a thing's being valuable *simpliciter*, or desirable, could be explained in terms of its being *fit to desire*. Alternatively, if being valuable isn't (always) a matter of being desirable, then FA theorists might instead explain facts about what's valuable in terms of facts about what's fitting to *value*, where this attitude of 'valuing' might consist variously in a plurality of pro-responses, including, *inter alia*, being glad that, respecting, taking pleasure in, and protecting.

The advantages of FA theories are various. For example, they seem to constitute an attractive middle ground between primitivism about evaluative properties and subjectivism about such properties (McHugh and Way 2016). Relatedly, FA theories of value *simpliciter* purport to "demystify" facts involving this property (Rabinowicz and Rønnow-Rasmussen 2004) and FA theories more generally stand to provide unifying explanations of the varieties of value (Schroeder 2010; Rowland 2019) and to explain the structure of, and various necessary connections within, certain parts of the normative and evaluative domain (Way 2013). But FA theories also face a number of outstanding issues, including concerns about circularity and extensional adequacy. This entry clarifies the nature of FA theories of value, highlights some important choice points for proponents of these theories, and surveys and catalogues the main advantages of the theories and the most pressing, outstanding issues that they face today.

1. Clarifying FA Theories: Some Choice Points

1.1 Metaphysical, Conceptual, or Semantic?

Early proponents of FA theory were sometimes unclear about whether their intent was to elucidate the natures of evaluative concepts, terms, or properties (e.g., Brentano 1889/2009; Broad 1930; Ewing 1939, 1947). In the last decade, however, FA theorists have tended to theorize mostly in a metaphysical mode, seeking to explain various evaluative properties and facts (McHugh and Way 2016; Howard 2019; Rowland 2019). This entry follows the current trend and characterizes FA theories as metaphysical theses. That being said, for many of the necessary equivalences entailed by FA theories of specific evaluative properties—e.g., between *being admirable* and being *fit to admire*—it seems plausible that their semantic and conceptual counterparts also hold with the corresponding kind of necessity (Brandt 1946; Kjellmer 1986). And some have recently argued that it's also true of conceptual necessity that something is generically valuable, or valuable *simpliciter*, if and only if it's fitting to value, or "worth valuing" (Svavarsdóttir 2014).

Many contemporary FA theorists, then, take their FA theories to provide constitutive or metaphysical accounts of evaluative facts and properties. But there are various ways to understand what a metaphysical account of a property is. Sometimes such accounts are said to reveal *essences* (Fine 1994), other times to state *property identities* (Jackson 1998), and still other times to specify *grounds* (Rosen 2015b). Common to contemporary FA theorists, however, is the claim that their metaphysical accounts of evaluative properties in terms of normative properties of human responses entail that facts involving the latter are *explanatorily prior* to facts involving the former, such that the former facts are always fully grounded in the latter. This is important for several reasons. First, it reveals a core commitment of FA theories, viz., that the evaluative facts they seek to explain obtain fully in virtue of normative facts about human responses, rather than the other way around. An FA theory of facts about what's *fearsome*, for instance, would entail that whenever something is fearsome, that's *fully because* it's fitting to fear. Second, this explanatory commitment would seem to rule out understanding FA theories as stating property identities, given the irreflexivity of the grounding relation (see the entry on metaphysical grounding).

The order of explanatory priority posited by FA theories is sometimes claimed to conflict with commonsense (Stratton-Lake and Hooker 2006): if something is fitting to value, for instance, isn't that precisely because it's valuable? In other words, doesn't the fittingness of valuing a valuable thing derive from the value of that thing rather than the reverse? And indeed, isn't the value of valuable things at least in principle ontically independent of any actual or possible human responses to those things? Such concerns are discussed further in sections 2.1 and 3.2.

1.2 FA Theory, Primitivism, and Subjectivism

The explanatory element of FA theories also distinguishes these theories from "primitivism" about their explananda. Whereas an FA theory of an evaluative property seeks to explain facts involving that property by appeal to normative facts about human responses, primitivism about the property holds that facts involving it are explanatorily basic or fundamental. A paradigmatic proponent of primitivism is G.E. Moore (1903), who held this view about the property of being (intrinsically) valuable *simpliciter*. Notably, Moorean primitivism about value *simpliciter* is in principle compatible with FA theories of various other evaluative properties. For example, one might favor an FA theory of *being pitiful*, but reject such a theory of being valuable (or vice versa). And more generally, FA theorists might be selective, for various reasons, about the evaluative properties they seek to explain. Possible reasons for such selectivity are discussed in section 3.

FA theories of evaluative properties can also be distinguished from "subjectivist" theories of those properties, which seek to explain evaluative properties entirely in terms of human responses that meet certain non-normative criteria (e.g., Lewis 1989). For example, a subjectivist might hold that whenever someone is admirable, that's fully because that person would be admired by anyone who knew all the facts, reflected adequately on them, etc. FA theories, in contrast, hold that what makes someone admirable is that admiring them meets some *normative* condition. What this condition is, precisely, is a matter of intramural debate (see section 1.3). If the condition in question can ultimately be explained in fully naturalistic terms (Smith 1994), then FA theories may also turn out to be subjectivist theories (McHugh and Way fc). However, FA theories *needn't* be subjectivist: instead, they might deny that the normative facts in terms of which they explain certain evaluative facts can themselves be explained in naturalistic terms.

1.3 Which Normative Property?

So far, this entry has used the word 'fittingness' to pick out the normative property that FA theories appeal to in their explanations of evaluative facts. The fittingness relation can be glossed as the relation in which a response stands to its object when the object merits, or is worthy of, that response. Thus to say that something is a fitting object of admiration is to say that it merits admiration, or is worthy of this attitude. Many FA theorists have held not only that fittingness is the right normative relation to appeal to in explaining the evaluative facts they seek to explain, but also that fittingness is explanatorily fundamental relative to the rest of the normative and evaluative domain, if not fundamental *simpliciter* (Broad 1930; Ewing 1939, 1947; Chappell 2012; McHugh and Way 2016; Howard 2019). But FA theorists needn't be committed to any of these claims. For instance, FA theorists might agree that evaluative facts should be explained in terms of facts about fitting responses, but reject that fittingness is explanatorily fundamental (normatively or otherwise). Alternatively, they might deny that evaluative facts should be explained in terms of facts about fitting responses in the first instance, opting instead to explain the relevant facts in terms of the satisfaction of some apparently distinct normative condition, which may or may not be explicable in naturalistic terms. Other candidates for the normative condition in question include a response's being such that one "ought" to have it (Sidgwick 1874), its being "correct" (Brentano 1889/2009), "justified" (Ewing 1959), "rational" (Brandt 1979; Anderson 1993; Smith 1994), "required" (Chisholm 1986; Lemos 1994), "warranted" (Gibbard 1990), or its being such that there's (some, sufficient, or decisive) "reason" for one to have it (Scanlon 1998; Schroeder 2010; Rowland 2019). The debate among FA theorists concerning which normative property to appeal to in their theories continues today. Presently, the two most popular options are fittingness and reasons. Some costs and benefits associated with the FA theorist's selection of either of these options are discussed in sections 2 and 3.

1.4 Fitting for Whom?

FA theories explain evaluative properties in terms of normative facts about responses that are somehow called for by things that have those properties. But *who* is to have the responses in question? This is undertheorized, but the answer seems to be: it depends. In particular, the answer seems to depend on the nature of the evaluative property that an FA theory seeks to explain. If the evaluative property is "agent-neutral", then everyone, or just anyone, is to have the response in terms of which FA theories aim to explain the property in question; if the property is "agent-relative", then at least someone, but not just

anyone, is to have the response. Whether an evaluative property is neutral or relative can be a matter of debate, but consider *awesomeness* as a plausible example of the former and *shamefulness* as an example of the latter. The awesome calls for everyone, or just anyone, to be in awe of it, but the same isn't so for shamefulness: not just anyone is to be ashamed of a shameful act; instead, a shameful act calls for shame only on the part of the person who performed it (or perhaps also on the part of those who are somehow closely related).

A further, related question is whether the normative property of responses that FA theories appeal to is (or should be) *objective*, such that all facts are in principle relevant to whether it obtains, or *perspectival*, such that only epistemically accessible or possessed facts are relevant. Here, too, it seems the FA theorist should say that it depends on the nature of the evaluative property that an FA theory aims to explain. For instance, since facts about whether someone is *lovable* or *blameworthy* don't seem to depend on anyone's epistemic position, such that someone could have these properties without anyone knowing it or otherwise being aware, FA theorists should say that such facts are to be explained by appeal to an objective normative property of responses. In contrast, since facts about the *curiousness* of a question or the *credibility* of a proposition can seem to depend on a subject's perspective, FA theorists should say that such facts are to be explained by appeal to a normative property that's similarly perspective dependent (Way 2020). This is a further way in which an evaluative property might be agent-relative, as opposed to agent-neutral: if it's perspectival, as opposed to objective. To be clear: this doesn't mean that FA theories are necessarily disunified, appealing to different normative properties depending on their explanandum; instead, FA theorists can seek to explain whatever objective normative property they appeal to in terms of its perspectival counterpart or vice versa, and claim that evaluative properties can ultimately be explained by appeal to whichever of these is more fundamental.

1.5 Buck Passing

Some suggest that FA theories date as far back as Kant (1785/1996) (see, e.g., Suikkanen 2009). More commonly, the progenitors of FA theory are taken to be Sidgwick (1874) and Brentano (1889/2009). Although many have discussed and defended FA theories in the meantime (for a nice potted history, see Rabinowicz and Rønnow-Rasmussen 2004) it's relatively uncontroversial that T.M. Scanlon (1998) is largely responsible for the contemporary revival of interest in FA theory. According to Scanlon's "buck-passing" account of value, "to call something valuable is to say that it has other properties that provide reasons for behaving in certain ways with regard to it" (96). There are several things to note about Scanlon's account. First, it purports to provide a theory specifically of being valuable *simpliciter*. Scanlon doesn't comment on whether he takes the form of his account to generalize to apply to more specific evaluative properties such as being lovable or delightful, though it's easy to see how it might. For example, we might claim that to call something delightful is to say that it has other properties that provide reasons for being delighted by it. Second, Scanlon isn't very specific about *how* we have reason to behave with regard to valuable things, i.e., about the specific human responses in terms of which he aims to explain facts about value *simpliciter*. This seems intentional. Scanlon suggests that "to claim that something is valuable (or that it is 'of value') is to claim that others also have reason to value it, as you do" (95), but holds that "valuing" might consist variously in a range of pro-responses, where the particular kind of pro-response to a valuable thing that we have reasons to have depends on, and so varies according to, the kind of valuable thing at issue. For example, on Scanlon's view, the pro-response(s) that we have reasons to have toward a valuable piece of music can differ from the pro-response that we have reasons to have toward a valuable friendship. And it can be matter of substantive normative dispute, he suggests, which pro-response we have reasons to have toward each. Thus, Scanlon writes: "understanding the value of something often involves not merely knowing that it is valuable or how valuable it is, but also how it is to be valued" (100). (For a possible difficulty with this part of Scanlon's view, see section 3.4.)

The final thing to note about Scanlon's account is its commitment to the claim that the property of being valuable doesn't itself provide a reason to value the things that have it. Rather, Scanlon claims, being valuable is the "purely formal, higher-order" property of having other properties that provide reasons for a pro-response. This is the feature of Scanlon's view that makes it a "buck-passing" account: the account passes the normative buck from the property of being valuable to lower-order properties of a thing that

provide reasons to respond to it favorably, where these lower-order properties correspond extensionally to the value-making features of the valuable thing in question. This feature of Scanlon's account has been called its "negative thesis" (Schroeder 2009). This is in contrast to its "positive thesis": that whenever something is valuable, that's fully because there are reasons to have a pro-response to it. The positive thesis is what makes Scanlon's theory an FA theory: it claims that certain evaluative facts—facts about value *simpliciter*—are always fully grounded in normative facts about certain human responses, viz., facts to the effect that the relevant responses are favored by reasons. As several authors have pointed out, however, Scanlon's negative thesis isn't entailed by the positive one and is otherwise inessential to FA theory (see esp. Schroeder 2009). The costs and benefits of the negative thesis are still being debated today (for some advantages, see Suikkanen 2009; for some costs, see Schroeder 2009 and Johnson King 2019). Perhaps the most promising, conciliatory approach is to claim that the property of being valuable does provide a reason to value the things that have it, but only a *derivative* reason which doesn't contribute any weight to the case for valuing the thing in question over and above that contributed by the non-derivative reasons provided by certain of the thing's lower-order properties (Schroeder 2009; McHugh and Way 2016; Rowland 2019). Regardless, since the negative thesis isn't plausibly essential to FA theory, whether to accept this or some other version of it is a further choice point for FA theorists.

2. Attractions

This section canvasses three main attractions of FA theories: their ability to explain various interesting connections between certain normative and evaluative properties (section 2.1); their promise to "demystify" various evaluative properties in various ways (section 2.2); and their potential to unify the evaluative properties they aim to explain by identifying what's common to them, while respecting and clarifying any important differences between them (section 2.3).

2.1 Normative-Evaluative Connections

A central attraction of FA theories is that they promise explanations of interesting connections between certain normative and evaluative properties (McHugh and Way 2016). For example, it seems difficult to deny that being valuable and being fitting to value necessarily co-vary with each other such that something has value if and only if it's fitting to value, or merits being valued (though see section 3.2 for discussion). And, as discussed in the introduction to this entry, similar equivalences also seem to hold between various specific evaluative properties, such as being admirable or delightful, and the fittingness of correspondingly specific types of human responses, e.g., admiring, or being delighted by. FA theories offer straightforward explanations of these equivalences, viz., that facts involving the evaluative property on the left-hand side are always fully grounded in facts involving the normative property of responses mentioned on the right, such that facts of the former kind obtain if and only if, and because, facts of the latter kind do.

As another example, consider that whenever something has value, it seems that there's at least some reason to value it. More precisely, it's plausible that, whenever p is a respect in which x is valuable, p is a reason to value x (Way 2013). For example, if the happiness it brings is a respect in which your friendship is valuable, then it's also a reason to value your friendship. An FA theory that explains facts about value in terms of facts about reasons for valuing is well-suited to explain this: What explains the fact that respects in which x is valuable are also reasons to value x is that whenever p is a respect in which x is valuable, that's fully because p is a reason to value x. And parallel explanations can be given of similarly plausible claims concerning the connections between various specific evaluative properties and reasons: respects in which x is admirable, lovable, or contemptible, for instance, are also reasons to admire, love, or contemn x, because facts involving the former properties are always fully grounded in facts involving the latter.

In principle, FA theorists who appeal to fittingness rather than reasons to explain evaluative facts could also explain the necessary connections between reasons and evaluative facts just mentioned. To do so, they'd need only to offer an account of the relationship between fittingness and reasons that, together with their FA theories, explains the connections in question (McHugh and Way 2016; Howard 2019). Likewise, FA

theorists who formulate their theories in terms of reasons rather than fittingness could try similarly to explain the former kinds of necessary connections, e.g., between being valuable and being fitting to value (Schroeder 2010). And more generally, regardless of which normative property an FA theorist appeals to in their explanations of evaluative facts, explanations of the above connections are in principle available to them.

In this respect, FA theories have an advantage over so-called "no-priority views" which hold that evaluative facts and normative facts of other kinds can't be fully, non-circularly explained by appeal to one another, such that neither kind of fact is more fundamental (Ross 1939; Wiggins 1987; McDowell 1998; Tappolet 2011). Indeed, at least barring their ability to provide naturalistic explanations of the normative and evaluative facts that figure in the above connections, it seems that no-priority theorists would need to take these connections as brute. In this way, FA theories have more explanatory power than no-priority views (Rowland 2019).

The potential to explain the above connections, however, isn't unique to FA theory. For example, consider a version of value primitivism, of the kind described in section 1.2, that would reverse the order of explanation posited by FA theories, explaining normative facts about human responses in terms of evaluative facts about the objects of those responses. A proponent of this kind of "object-focused" primitivism might hold that whenever something is fitting to value that's fully because that thing is valuable, and that this is what explains why something has value if and only if it's fitting to value. And they might propose similarly to explain the parallel equivalences between the fittingness of various other types of response and other evaluative properties. Likewise, an object-focused view that claims that facts about reasons to value things are always fully grounded in the value of those things might explain why there are always reasons to value valuable things (Orsi 2013b). Object-focused views thus seem able to explain the normative-evaluative connections noted above. Moreover, as was mentioned in section 1.1, some find the order of explanation posited by this kind of primitivism more intuitively plausible than that claimed by FA theory (Hooker and Stratton-Lake 2006). Still, FA theories do seem to have important advantages over object-focused views, in explaining the connections in question.

For one, FA theory is far more parsimonious. Suppose we explain normative facts about human responses (whether they're facts about the fittingness of those responses or facts about reasons for them) in terms of evaluative facts about the objects of those responses, as object-focused views suggest. So, for example, we explain facts about fitting admiration in terms of facts about the admirability of admiration's object and facts about fitting love in terms of facts about the lovability of love's object. Assuming these evaluative facts couldn't then be explained in terms of normative or evaluative facts of some other kind, it seems that we'd thus be left with a (massive) plurality of different kinds of evaluative fact, each of which would be fundamental at least relative to the rest of the normative and evaluative domain. Moreover, assuming that different kinds of normative relations are to be individuated according to the grounds of facts involving those relations (Leary 2020), an object-focused view would entail, e.g., that the fittingness of admiration is a different kind of normative relation than the fittingness of love, insofar as facts involving each are fully grounded in different kinds of evaluative fact. Thus, object-focused views would need to posit not only a plurality of different kinds of fundamental evaluative facts, but also a plurality of different kinds of fittingness (or reasons), each corresponding to the specific kind of evaluative fact in terms of which they explain facts involving the relation in question. Object-focused views in this way disunify the normative relation(s) they purport to explain.

Alternatively, if, as the FA theorist proposes, we explain evaluative facts ultimately in terms of facts involving a normative relation of a single, unified kind, then we could in principle take only facts involving that relation to be fundamental relative to the rest of the normative and evaluative domain. Indeed, even if we didn't take facts involving the relevant relation to be fundamental and instead explained them in terms of facts involving some other normative property of human responses, we'd still have to posit fewer kinds of fundamental facts as compared to object-focused primitivism. Thus, FA theories are more parsimonious than object-focused views.

A second version of value primitivism, however, fares better along this dimension. On this view—call it "response-focused" primitivism—normative facts about human responses are to be explained not in terms of evaluative facts about the *objects* of those responses, but rather in terms of evaluative facts about the responses themselves (Moore 1903). In particular, this view holds that normative facts about our responses are always fully grounded in facts to the effect that it's somehow valuable (*simpliciter*) to have those responses. Similar to object-focused views, response-focused views also seem able to explain the normative-evaluative connections noted above. For example, response-focused primitivists might claim that something is valuable if and only if it's fitting to value because whenever something is fitting to value, that's fully because it's intrinsically valuable to value it; and, as a substantive matter, something is intrinsically valuable to value if and only if it's valuable (Hurka 2000). And the response-focused theorist might similarly explain the parallel equivalences between various other evaluative properties and the fittingness of certain other responses, arguing that it's always intrinsically valuable, e.g., to admire the admirable, love the lovable, despise the despicable, etc. Likewise, response-focused views might claim to explain why there are always *reasons* to value valuable things: they might hold that whenever there are reasons to value something, that's fully because valuing it is valuable and, as a substantive matter, if something is valuable, then it's valuable to value.

In contrast to object-focused views, response-focused views needn't be less parsimonious than FA theory. In principle, a response-focused primitivist could hold that all other normative and evaluative facts are to be explained ultimately in terms of facts about the value *simpliciter* of various types of human response (or states of affairs, or whatever they take to be the fundamental bearers of value) (Moore 1903). However, response-focused views do face extensional and explanatory worries that bear on the plausibility of their possible explanations of the normative-evaluative connections under discussion (Way 2013; Howard 2018). The main extensional worry concerns the response-focused view's claim that, whenever a response is fitting, it's valuable to make that response. Perhaps it's somehow valuable to value what's valuable, but is it also valuable to be disgusted by the disgusting, to envy the enviable, and to be bored by the boring? If not, then response-focused views would seem unable to explain, e.g., why it's fitting to be disgusted by something if and only if it's disgusting. And, for similar reasons, they'd also be unable to explain, e.g., why there are reasons to be disgusted by the disgusting, to be bored by the boring, etc.

FA theories thus look to have some important *prima facie* advantages over at least many of their competitors in explaining the normative-evaluative connections noted above. Of course, none of these advantages is necessarily decisive with respect to overall theory choice and there is, in any case, room for rivals of FA theory to dispute them. For example, an opponent of FA theory might suggest that the outstanding issues for FA theory that we'll discuss in section 3 are at least as serious as the worries for primitivism raised above. Nevertheless, it remains an attraction of FA theory that it seems well placed to explain the normative-evaluative connections in question.

2.2 Value "Demystified"

A second attraction often claimed for FA theory is its potential to "demystify" value in various ways. Rabinowicz and Rønnow-Rasmussen (2004) suggest, for example, that FA theory "removes the air of mystery from the normative 'compellingness' of values" (391). Their thought is that it's no mystery why facts about the value of things should have normative implications for our actual and possible responses to those things if such facts are always fully grounded in, and so imply, normative facts about human responses. The idea, in effect, is that FA theories stand to explain the normative-evaluative connections discussed in section 2.1. As was noted there, it's at least possible for primitivists about evaluative properties to explain these connections, too. But there are other interesting ways in FA theories have been claimed to demystify value and, in particular, value *simpliciter*, which can't even in principle be matched by primitivists about this property.

Perhaps the most important such way of demystifying value is by answering skepticism about it. In an influential paper, Peter Geach (1965) argues that there's no such thing as being good or valuable *simpliciter*. Although Geach doesn't remark specifically on the existence of specific evaluative properties,

such as being admirable or delightful, he does argue, roughly, that the only intelligible kind of "goodness" there is, is "attributive goodness", i.e., the kind of goodness that a thing has when it's good as the kind of thing that it is. Thus, for Geach, while it makes sense to claim that things can be good *as* the kind of thing they are—e.g., that there can be good toasters, baristas, and bombs—there's no sense to be made of claims to the effect that something is *just plain good*, i.e., valuable in a way that isn't in the relevant way relativized to a particular kind of thing. Citing Geach, Philippa Foot (1985) and Judith Jarvis Thomson (1993, 1997) also marshal versions of this idea against traditional forms of act consequentialism, which, on a natural characterization, aim to explain deontic properties of acts ultimately in terms of facts about the (comparative) value *simpliciter* of their outcomes. Although Foot and Thomson disagree with Geach that *all* value is attributive (acknowledging the existence and coherence of a distinct notion of being valuable *for*, for instance), they at least agree that there's no such thing as being just plain valuable, or valuable *simpliciter*. This conclusion is independently interesting, but, if true, it would also have devastating implications for traditional forms of consequentialism (not to mention for primitivism about value *simpliciter* of the kind discussed in sections 1.2 and 2.1).

However, as several authors have noted, FA theories seem to offer a promising response to this kind of skepticism (Zimmerman 2001; Nye, Plunkett, and Ku 2015; Rowland 2016; McHugh and Way 2016). In particular, FA theorists might hold that facts about value *simpliciter* can be explained fully in terms of certain kinds of normative facts about certain kinds of valuing responses. Since skeptics about value *simpliciter* aren't similarly skeptical about the kinds of normative facts in terms of which FA theorists might propose to explain facts involving this property, the FA theorist's account of value *simpliciter* would thus seem to render facts involving the property intelligible, even by the skeptic's own lights. In this way, FA theorists can answer skeptics about value *simpliciter*.

And skepticism about value *simpliciter* isn't the only kind of value skepticism that FA theories stand to answer. In the last several decades, a version of act consequentialism known as "agent-relative" consequentialism has come to prominence, promising to account for a range of normative phenomena often held to be incompatible with traditional formulations of the view, e.g., special obligations, permissions, and deontic constraints (Dreier 1993; Portmore 2005; Smith 2003). This kind of consequentialism can be developed in either of two ways. On the first way, agent-relative consequentialism posits a distinctive evaluative relation, viz., that of being valuable *relative to* an agent, where this 'agent-relative value' is meant to be distinct both from value *simpliciter*, which is characteristically held to be agent neutral, and from the relation of being valuable *for* some person or thing (see esp. Portmore 2005). Agent-relative consequentialists who posit this distinctive relation thus claim that which outcome is best *relative to an agent* might differ from the outcome that's best *simpliciter*, or best for the agent. And, at least sometimes, what an agent ought to do on this view is what would produce the outcome that's best relative to them. Paired with the right substantive assumptions about how to rank outcomes relative to particular agents, this apparatus purports to allow agent-relative consequentialists to capture the normative phenomena noted above. The second way of developing the view works similarly except that, rather than positing a distinctive notion of agent-relative value, on top of value *simpliciter* and the relation of being valuable for, this variant of the view claims that, in fact, there's just a *single* kind of value, whose nature is fundamentally agent-relative, and which agents ought always to maximize (Smith 2009).

However, in an influential paper, Mark Schroeder (2007b) questions the intelligibility of the kind of agent-relative value to which these versions of consequentialism appeal. And many others have also claimed, similarly, that they lack any independent, pre-theoretical grip on the notion of agent-relative value (Dancy 2000; Hurley 2017; Maguire 2016; Zimmerman 2011). In response, agent-relative consequentialists (of both kinds) have turned to FA theory, claiming that the relevant evaluative relation can be rendered intelligible by explicating it in terms of certain normative properties of preference or desire (see esp. Portmore 2011 and Smith 2009). For example, the agent-relative consequentialist might claim that the agent-relative value of an outcome can be explicated in terms of agent-relative reasons for preferring it to its alternatives. They might then claim what agents ought to do is to bring about the outcomes they ought to prefer, all things considered, with the stipulation that agent-relative reasons for preference can be relevant to which outcomes agents ought to prefer. The details are important here, and certain ways of developing this view seem less tenable than others (see, e.g., Schroeder 2007, 292-94). Still, FA theories do seem to

offer agent-relative consequentialists the most promising line of response to their skeptical critics. (For more on this application of FA theory, see the entry on value theory).

2.3 Unifying the Evaluative

A third attraction of FA theory is its potential to unify the evaluative properties it aims to explain, by identifying what's common to those properties, while also clarifying important differences between them. As an initial example, consider the various ways of being valuable *simpliciter*. A thing might be valuable instrumentally, noninstrumentally, intrinsically, extrinsically, in some respect, or overall. An FA theory of being valuable *simpliciter* can be applied to provide unifying explanations of all these ways of having this property (McHugh and Way 2016; Rowland 2019). Perhaps the most popular way to do this is to appeal to different ways of *valuing*. For instance, an FA theorist might hold that whenever something is noninstrumentally valuable that's fully because it's fitting to value *for its own sake*, whereas when something is valuable instrumentally, that's fully because it's fitting to value ultimately *for the sake of something else* (viz., something that's fitting to value for its own sake). Similarly, FA theorists might claim that whenever something has intrinsic (extrinsic) value, that's fully because it's fitting to value that thing intrinsically (extrinsically), i.e., for its intrinsic (extrinsic) properties. Likewise, the FA theorist might explain facts about a thing being valuable in some respect (overall) in terms of facts to the effect that it's fitting to value that thing in that respect (overall). And these accounts can of course be combined to explain any way of being valuable that consists in a complex of some possible subset of any of the above ways of possessing this property. For instance, an FA theorist might claim that whenever something is noninstrumentally, extrinsically valuable in some respect, that's fully because it's fitting to value that thing for its own sake, for its extrinsic properties, in the relevant respect. And there are other possibilities for the FA theorist here, beyond appealing to different ways of valuing. For example, an FA theorist might propose to explain the difference between something being valuable in a respect as opposed to being valuable overall not by appeal to differences in the ways of valuing that figure in their accounts of these properties, but rather by appeal to differences in the normative properties that their accounts deploy. For instance, Rowland (2019) suggests that whenever something is valuable in a respect, that's fully because there's a *pro tanto* reason to value it, whereas when something is valuable overall, that's fully because there's *sufficient reason* to value it. Which approach is more plausible is a matter of intramural debate.

In addition to offering unifying explanations of all the ways of being valuable *simpliciter*, FA theories also promise to unify other varieties of value. For example, just as we might offer FA theories of being valuable *simpliciter*, we might also offer them of the properties of being *attributively* valuable and of being valuable *for* some person or thing (Schroeder 2010; Skorupski 2010; McHugh and Way 2016; Rowland 2019). For example, McHugh and Way (2016) suggest as a first pass that whenever something, *x*, is attributively valuable, or valuable as a kind of thing *K*, that's fully because it's fitting for anyone for whom it's fitting to want a *K* to value *x*. And according to Rowland (2019), whenever something, *x*, is valuable for some person or thing, *S*, that's fully because there are reasons for anyone who has reasons to care about *S* to value *x*, because they have reasons to care about *S*. And the prospects seem at least equally good for a fittingness-based FA theory to explain the relation of being valuable for, and for a reasons-based FA theory to explain the property of being attributively valuable. So, regardless of which normative property of responses FA theories appeal to, they have the potential to unify these different varieties of value. And since it's plausible that the different varieties of value must have something in common in virtue of which they all qualify as varieties of *value*, it's an advantage of FA theories that they're able to explain what this is—viz., that each can be fully explained in terms of facts involving normative properties of responses to be had by certain sets of agents—while also respecting the real, important differences between the kinds of value in question.

And FA theories also seem able, similarly, to unify the variety of specific evaluative properties that populate the evaluative domain, e.g., that of being admirable, delightful, praiseworthy, etc. For assuming that, for each such property, facts involving it can be explained fully in terms of facts involving a certain normative property of a certain type of response, FA theorists can claim that this is what all such properties have in common, while also properly distinguishing among them according to the particular type of response that they appeal to in their explanation of each. And, according to some, there are likewise prospects for applying FA theories to so-called "thick" evaluative properties, such as that of being

courageous and kind (for extensive discussion, see Rowland 2019, ch. 9). If so, FA theories stand to provide unifying explanations not just of proper subsets of evaluative properties, but of *all* such properties. Indeed, the most ambitious FA theorists might claim that all evaluative properties have in common that they can each be explained fully in terms of a certain normative property of a certain kind of human response.

3. Outstanding Issues

Despite the attractions of FA theories canvassed so far, these theories also face several outstanding issues which challenge both their extensional and explanatory components. This section covers four outstanding issues for FA theory. The first three pose extensional challenges: the Wrong Kind of Reason problem (section 3.1); the Wrong Kind of Value problem (section 3.2); and the problem of partiality (section 3.3). The fourth and final issue challenges the explanatory component of FA theory, alleging that at least certain FA theories may be viciously circular (section 3.4).

3.1 The Wrong Kind of Reason Problem

Perhaps the most famous challenge to FA theories is what's known as the Wrong Kind of Reason problem (WKR problem). At a first pass, the WKR problem is the problem that, at least sometimes, it seems that what's valuable doesn't correspond to what there's reason to value. For example, consider an FA theory like Scanlon's buck-passing view from section 1.5, which entails that something is valuable if and only if (and because) there are reasons for just anyone to value it. The trouble is that, intuitively, there can be reasons to value things that have no value as well as reasons not to value things that do have value. To adapt an example from Roger Crisp (2000), suppose an evil demon will kill you unless you value a cup of mud. This fact seems like a reason to value the cup, but the cup, being a cup of mud, has no value. Hence, an FA theory like Scanlon's buck-passing view seems sometimes to predict that the object of a valuing attitude has value when, intuitively, it doesn't. Similarly, consider an FA theory that explains facts about what's valuable in terms of facts about what there's *sufficient* reason to value. Now suppose the demon will kill you if you value something of value. Here, we might think you lack sufficient reason to value the valuable thing. If so, then this version of FA theory also seems subject to counterexample. This class of counterexamples is known as the "Wrong Kind of Reason problem" since the reasons that figure in the examples—reasons that (dis)favor a valuing attitude without making the attitude's object (dis)valuable—are of the "wrong kind" to figure in FA theories.

The WKR problem generalizes in two important ways. First, it seems to arise not only for certain FA theories of value *simpliciter*, but also for certain FA theories of more specific evaluative properties. For example, Crisp's demon might equally threaten harm unless you admire someone who's not admirable, despise someone who's not despicable, love someone who's not lovable, etc. Second, in addition to arising for versions of FA theory that explain evaluative facts in terms of facts about *reasons* for certain responses, the WKR problem also arises for FA theories that try to explain evaluative facts by appeal to deontic facts of various other kinds—for example, by appeal to facts about which responses we ought to or may have. Indeed, just as the demon's threat might give you a reason to value the mud, it might also make it the case that you *should* or that you're *required* to value it, or that valuing it would be *justified*. The question of whether the WKR problem generalizes to apply to *all* versions of FA theory is discussed in detail below.

As is always the case when a theory is confronted with putative counterexamples, two broad responses are available: proponents of the theory might either reject the counterexamples or concede them, but try somehow to revise the theory so as to avoid them. Both kinds of response have been explored by proponents of FA theory. The first kind of response amounts to skepticism about wrong-kind reasons, the view that such "reasons" aren't really reasons at all and hence aren't counterexamples to FA theory. On this view, the demon's threat doesn't give you a reason to value the mud (or make it the case that you should or may value it); instead, it at best gives you a reason to *want* to value the mud, or to bring it about that you do if you can. Thus, FA theories that explain facts about value by appeal to facts about reasons (or oughts,

permissions, etc.) don't falsely predict, e.g., that the mud has value. Hence, for WKR skeptics, putative WKRs pose no problem for FA theories (Parfit 2011; Skorupski 2010; Way 2012).

The success of this response is a matter of debate. According to many, the claim that WKRs for attitudes aren't reasons for those attitudes seems implausible on substantive, first-order grounds. After all, the fact that the demon will kill you unless you value the mud does seem to speak in favor of valuing it, and it also seems this fact could contribute to making it the case that you ought to value the mud, or that you'd be justified in valuing it. And, for various reasons, many have found unsatisfying the skeptic's attempt to save the phenomenon by claiming that, although the demon's threat gives you no reason to value the mud, it does give you a reason to want or to try to bring it about that you do (see, e.g., D'Arms and Jacobson 2000b, Rabinowicz and Rønnow-Rasmussen 2004, Danielsson and Olson 2007, Reisner 2009, and Rosen 2015a).

WKR skeptics thus need to argue that WKRs aren't reasons for the attitudes they seem to favor and to explain why they're not (McHugh and Way 2016). One popular proposed explanation of why WKRs aren't reasons appeals to a response-condition on reasons, according to which a consideration p can be a reason for you to A only if you could A for the reason that p (Kelly 2002; Shah 2006; Parfit 2011). However, proponents of FA theories that explain facts about value in terms of facts about reasons seem committed to there being reasons that violate such a condition (Way 2012). To adapt a case from Schroeder (2007a), suppose Nate loves successful surprise parties thrown in his honor but hates unsuccessful ones. Plausibly, if there's a surprise party for Nate next door, then there's value in his going and hence, on a reasons-based FA theory, a reason for him to want to go. However, since Nate hates surprise parties he expects, this isn't a reason for which he could want to go: "for as soon as you tell him about it, [the reason] will go away" (Schroeder 2007a, 165). Hence, reasons-based FA theories seem committed to there being reasons that violate a response condition. Thus, an appeal to this condition to explain why WKRs aren't reasons looks untenable as part of a reply to the WKR problem as it applies to these theories.

The second strategy for responding to the WKR problem, in contrast, accepts that WKRs are reasons. Proponents of this second strategy concede the counterexamples that constitute the WKR problem and try to revise FA theories in light of them. The idea is to distinguish between the "right" and "wrong" kinds of reasons to figure in FA theories, and to reformulate the theories such that they reference only reasons of the "right" kind. The distinction between right-kind reasons (RKRs) and WKRs for valuing attitudes can of course be easily drawn by appeal to the property of being valuable *simpliciter*: RKRs to value x are reasons to value x that also make x somehow valuable, whereas WKRs to value x are reasons to value x that don't also make x valuable. Likewise, RKRs to admire x are reasons to admire x that also make x admirable, RKRs to despise x are reasons to despise x that also make x despicable, etc. But such accounts of the RKR/WKR distinction are of course unavailable to FA theorists, at least insofar as they aspire to provide non-circular accounts of the evaluative properties that are the targets of their theories.

Over the last two decades, many attempts have been made to the draw the RKR/WKR distinction in a way that doesn't render FA theories circular. An early proposal held that RKRs are given by facts about the objects of the attitudes they favor, whereas WKRs are given by facts about the favored mental states themselves (Stratton-Lake 2005; Piller 2006; Parfit 2011). It's now widely held, however, that the object-/state-given theory is extensionally inadequate—that some WKRs are object-given (Rabinowicz and Rønnow-Rasmussen 2004) and that some RKRs are state-given (Schroeder 2012). Another view—the "alethic view"—holds that right- but not wrong-kind reasons to A are reasons to A which bear on the truth of A's content (D'Arms and Jacobson 2000a; Tappolet 2011; Rosen 2015). But this proposal has also been argued to face significant difficulties. First, it seems to require taking on substantial and contentious commitments about the contents of various types of attitude in order to guarantee the account's extensional correctness (McHugh and Way 2016; Schroeder 2010); and second, these commitments may ultimately render FA theories circular, if it's held that certain of these attitudes have evaluative contents—e.g., if valuing x is said essentially to involve thinking that x is somehow valuable (Ross 1939).

In addition to these, many other accounts of the RKR/WKR distinction have been offered and criticized (see, e.g., the proposals in Schroeder 2010 and Rowland 2019 and the criticisms of each, respectively, in Sharadin 2013 and McHugh and Way fc). Common to many proposed ways of drawing the RKR/WKR

distinction is a commitment to the idea that the distinction can be drawn in terms of reasons and certain non-normative properties alone. This commitment is common, for instance, to all of the proposals mentioned so far. But none of these proposals has gained wide acceptance. Partly in light of this, a number of authors have opted to avoid this commitment, and instead to offer an account of the RKR/WKR distinction that takes FA theories back to their earliest roots. On this approach, the RKR/WKR distinction can be drawn in terms of the normative notion of fittingness. Roughly, RKRs for attitudes are considerations that contribute to the fittingness of those attitudes, whereas WKRs for attitudes are reasons for attitudes that lack this feature. On this proposal, an FA theory of value *simpliciter* amounts to the thesis that facts involving value *simpliciter* are always fully and ultimately grounded in facts about the *fittingness* of certain valuing responses, rather than in facts about *reasons* for those responses.

Proponents of this position argue that the WKR problem poses no problem for FA theories of value so understood (Danielsson and Olson 2007; Chappell 2012; Sharadin 2015; McHugh and Way 2016; Howard 2019). They claim that while facts like the demon's threat—WKRs—may provide reasons for the attitudes they support, or even make it the case that you ought to have those attitudes, such facts are irrelevant to whether those attitudes are fitting. For example, even if the fact of the demon's threat gives you a reason to value a cup of mud, or makes it the case that you ought to have this attitude, this fact isn't one in virtue of which the mud *merits* being valued, or is worthy of this attitude; hence, this fact fails to make the attitude fitting (see the characterization of fittingness, in terms of merit and worthiness, offered in section 1.3). The judgment that WKRs such as the demon's threat are irrelevant to the fittingness of the attitudes they favor is widely shared. In addition to the authors cited above, see, *inter alia*, Broad 1930; Brandt 1946; Ewing 1939, 1948; Wiggins 1987; and McDowell 1998. Hence, it's been argued, drawing the RKR/WKR distinction in terms of fittingness, and so formulating FA theory ultimately in terms of fittingness rather than in terms of some (other) deontic notion, seems to offer a promising solution to the WKR problem for FA theories of value *simpliciter*. Further, this strategy promises to generalize to address the problem as it occurs for FA theories of other, specific evaluative properties as well.

At least on the face of things, then, FA theories that explain evaluative facts ultimately in terms of facts about the fittingness of certain types of response seem to have an advantage when it comes to answering the WKR problem. Still, some have raised concerns with the proposed appeal to fittingness in addressing this problem. For example, some authors claim to lack a pre-theoretical grip on the notion of fittingness, and thus find explanations of evaluative facts in terms of facts involving the fittingness relation uninformative (Rabinowicz and Rønnow-Rasmussen 2004). And others have argued, on various grounds, that facts involving the fittingness of human responses will ultimately need to be explained in terms of reasons (of a certain kind) for those responses (Schroeder 2010, 2020). A fully satisfying answer to the WKR problem that appeals to the notion of fittingness will thus need to address these challenges.

3.2 The Wrong Kind of Value Problem

The Wrong Kind of Value problem (WKV problem) also poses a difficulty for the extensional component of at least certain FA theories. It alleges that there can be cases in which something has value, but it's not the case that any valuing response, on the part of anyone, is broadly speaking called for by that thing—i.e., the thing isn't fitting for anyone to value, there's no reason for anyone to value it, it's not the case that anyone ought to value it, etc. In short: there are cases where facts about value *can't* correspond to normative facts to the effect that some valuing response is called for, since in some cases the former sort of fact can obtain without the latter.

To illustrate, consider an example, the spirit of which is due to Krister Bykvist (2009), in which there are happy egrets, but no past, present, or future agents, i.e., beings capable of making a valuing response. Plausibly, this world, or at least the feature of it that consists in the egrets' pleasure, has value. But since there are no beings in this world capable of valuing the egrets' pleasure, the egrets' pleasure can't be valued and hence a valuing response to it can't be fitting or otherwise called for. Thus, the value of the egrets' happiness doesn't correspond extensionally to a normative fact about a valuing response. Hence, FA theories must be extensionally false.

Another kind of example is due to Andrew Reisner (2015). In this kind of example, the value of something in the actual world depends on whether anyone in the actual world values that thing. If anyone in the actual world does value the thing, then the thing is disvaluable, or bad. However, if no one values it, then it's valuable. (There are various ways of filling out this kind of case, but these are its important structural features; for Reisner's preferred version, see ibid., 467-8.) In this kind of case, valuing something valuable extinguishes its value. Reisner claims that it therefore can't be fitting to value the valuable thing; hence, FA theories are extensionally inadequate.

Defenders of FA theory have responded to these cases in various ways (Orsi 2013a; Way 2013; Elliott 2017; Rowland 2019). Perhaps the most popular response to each is one that purports to answer both. This response agrees that it's not fitting (or otherwise called for) for anyone in the happy egret world to value the egrets' pleasure (since there are no beings capable of valuing in this world) and that it's not fitting for anyone in the world Reisner describes to value the thing whose value would be extinguished if anyone in that world were to value it. But, the response goes, it's nonetheless fitting for beings in *other* possible worlds—us, for example—to value the egrets' pleasure (so far as it goes) and to value the valuable thing whose value would be eliminated were to it be valued by its worldmates (Orsi 2013a; Way 2013; Rowland 2019). Hence, in both kinds of case, the value at stake can still be said to correspond extensionally to normative facts about a certain valuing response, albeit a response to be made by beings of other worlds.

Reisner (2015) anticipates and criticizes the above response on various grounds, though see Rowland (2019, ch. 7) for an extended reply. A different response, also offered by Rowland, suggests that FA theorists might state their theory counterfactually, claiming (roughly) that whenever something has value, that's fully because valuing it *would* be called for were there any beings capable of valuing it around. On this formulation of FA theory, something has value if, in the nearest world where there are beings capable of valuing it, their valuing it is called for. Since a valuing response to the egrets' pleasure is evidently called for in the nearest world where there are beings capable of valuing it, this formulation of the theory thus predicts that the birds' pleasure has value. Thus, this formulation of FA theory seems at least to avoid Bykvist's kind of counterexample. However, it doesn't also obviously avoid Reisner's. In Reisner's case, unlike Bykvist's, the nearest world where there are beings capable of valuing the valuable thing in question is the *actual* world. But recall: if beings in the actual world were to value the valuable thing, the thing wouldn't be valuable. And according to Reisner, this means it can't be fitting or otherwise called for, for those beings to value the valuable thing. Reisner here relies on a principle according to which it can be fitting for an agent S to value a thing x only if x would be fitting to value, conditional on S's valuing it. If this principle is true, then the counterfactual proposal fails to address Reisner's case; but see Rowland (ibid.) for arguments that it's false.

Which of these responses to the WKV problem is preferable, and whether either is ultimately acceptable, remain open questions, and so constitute avenues for further research. Two final points are worth emphasis. First, the force of the WKV problem, unlike that of the WKR problem, seems to depend not at all on which normative property of responses that FA theories appeal to in their explanations of evaluative facts. In particular, the problem seems to apply with equal force to FA theories that appeal to the *fittingness* of human responses and to those that appeal to *reasons* for those responses.[1] Hence, the WKV problem may apply more widely than the WKR problem. But, secondly, the WKV problem seems in another way narrower than the WKR problem. For while the WKR problem generalizes to apply not only to FA theories of generic value, or value *simpliciter*, but also to FA theories of specific evaluative properties, such as those of being shameful or disgusting, it's less clear that the WKV problem also generalizes in this way. For notice that the WKV problem illustrates and supports the worry, mentioned in section 1.1, that facts about value *simpliciter* may seem ontically independent of (normative) facts about actual or possible human responses. But it seems less clear that there's similar temptation to hold, for instance, that facts about what's *shameful* or *disgusting* could be ontically independent of (normative) facts about the human

---

[1] That being said, certain proposed solutions to the problem, e.g., that in Elliott 2017, may work better for reasons-based FA theories than for fit-based ones (or vice versa). Thanks to Jonathan Way for this point.

responses of shame or disgust. Perhaps there could be value in a world without valuers, but could something be shameful in a world without shame? If not, then the WKV problem may be a problem for some, but not all, FA theories, depending on the evaluative properties the theories target. At best, then, the WKV problem wouldn't be a reason to reject FA theory *tout court*, but only to be more selective about which FA theories we accept.

3.3 The Problem of Partiality

In addition to offering accounts of monadic evaluative properties, such as those of being lovable or valuable, FA theories of dyadic, comparative evaluative properties, such as those of being more valuable or *better than*, can also be given. For example, FA theorists might claim that whenever x is better than y, that's fully because it's fitting to value x more than y, or to prefer x to y. Likewise, they might hold that whenever x is more lovable than y, that's fully because x is fitting to love more than y. Indeed, according to some, FA theorists should take monadic properties like that of being valuable or lovable to be explanatorily posterior to their comparative counterparts (Schroeder 2010). But the problem of partiality poses an extensional worry for FA theories of at least certain comparative evaluative properties, for example, that of being better than. The problem, roughly, is that in some cases where y is no better than x—or indeed, where x is worse than y—it can seem fitting to be partial to x, or to prefer x to y. For example, suppose My Pal and Stranger are both up for the job. And suppose too that, except for the fact that My Pal is my pal and that Stranger is a stranger, all else is equal—they'd perform equally well in the role, they'd be equally happy if they got it, equally disappointed if they didn't, etc. Plausibly, then, My Pal's getting the job is no better than Stranger's getting it and vice versa—the two outcomes seem equally good. Still, it seems fitting for me to prefer that My Pal gets the job. Hence, it seems fitting for me to prefer one outcome to another when the outcomes are equally good. This is the kind of example that constitutes the partiality problem; and instances are easily multiplied.

The partiality problem was noticed and discussed by one of the progenitors of FA theory, A.C. Ewing (1939), but has only recently come to occupy the attention of contemporary authors (e.g., Bykvist 2009; Olson 2009; Lemos 2011; Zimmerman 2011; and Orsi 2013). Importantly, the problem seems to apply only to FA theories of comparative evaluative facts that propose to explain the relevant facts in terms of normative facts that call for just anyone to have a certain response. For example, that it's fitting for me to prefer that My Pal gets the job seems problematic only for FA theories of betterness that propose to explain facts about what's better than what by appeal to facts about what's fitting for just anyone to prefer (or to value more). This way of explaining the property of being better than is a natural way of capturing the popular and plausible idea that betterness is an agent-neutral evaluative property (see section 1.4). However, if we instead hold that betterness is fundamentally agent-relative, in the manner of Smith (2009), then we might reject the idea that facts about it are to be explained always in terms of normative facts about what's fitting for just anyone to prefer (as Smith himself does). This would be one way of avoiding the partiality problem as it applies to FA theories of better than, though the view about betterness on which it rests is controversial; by far, the majority view is that the property is agent-neutral.

A second way of dissolving the partiality problem is suggested by Jonas Olson (2009). On Olson's view, the partiality problem is structurally the same as, and indeed a special case of, the WKR problem (see section 3.1): the fact that My Pal is my pal gives me (some, sufficient, or decisive) reason to prefer the outcome where they get the job to the outcome where Stranger does, but it doesn't make the former outcome any better than the latter. Hence, according to Olson, a solution to the WKR problem should also suffice to solve the problem of partiality. On Olson's favored solution to the WKR problem, FA theorists should explain evaluative facts in terms of facts about fittingness rather than reasons. Olson suggests that while the fact that My Pal is my pal might give me *reason* to prefer that they get the job, this fact doesn't also make this preference *fitting*. Hence, FA theories that explain betterness in terms of what's fitting for all agents to prefer, rather than in terms of what there's reason for all agents to prefer, can avoid the problem of partiality.

However, there are two reasons to doubt Olson's diagnosis of the problem of partiality as an instance of the WKR problem. First, whatever we might think about the plausibility of WKR skepticism as a solution to the WKR problem, this view seems much less plausible as a solution to the problem of partiality. Even those inclined to deny that the demon's threat gives you a reason to value a cup of mud would likely find it hard to deny that My Pal's being my pal gives me a reason to prefer that they get the job. Further, it's not obviously incoherent for the WKR skeptic to deny that the former fact is a reason while accepting that the latter one is. Thus, the fact that WKR skepticism seems at least more viable as a solution to the WKR problem than it does as a solution to the problem of partiality suggests that these problems aren't the same.

Second, as indicated above, it seems quite plausible not only that I have reason to prefer that My Pal gets the job, but that, contrary to Olson, this preference is fitting. This, too, distinguishes the examples that constitute the problem of partiality from those that constitute the WKR problem. As discussed in section 3.1, the demon's threat seems not to make it fitting to the value the mud—it's not a fact in virtue of which the mud *merits* being valued. By contrast, the fact that My Pal is my pal does seem to merit my preferring that they get the job; hence, this preference is fitting.

Both of these considerations suggest that the problem of partiality isn't an instance of the WKR problem. But they also make clear that, like the WKV problem (section 3.2), the problem of partiality can pose a problem for certain FA theories regardless of the normative property they appeal to: Just as the fact that My Pal is my pal gives me a reason to prefer that they get the job, or perhaps makes it the case that I should have this preference, this fact also makes this preference fitting. Hence, the problem of partiality is in this way a rather general problem for FA theories, specifically for those that aspire to explain certain comparative evaluative facts.

Beyond Olson's proposal, various other solutions to the partiality problem have been suggested and debated (see, e.g., Lemos 2011; Zimmerman 2011; Orsi 2013a; and Nye, Plunkett, and Ku 2015). However, it's fair to say that, so far, none has gained wide acceptance. Thus, the partiality problem, perhaps more so than the WKR problem or the WKV problem, remains an outstanding and pressing issue for FA theory.

3.4 Circularity

As noted in the introduction to this entry, some FA theorists identify what's valuable *simpliciter* with what's desirable, and explain facts about what's desirable in terms of facts about what's fitting to desire. Other FA theorists, however, reject that the valuable can be identified with the desirable, often on the grounds that some valuable things seem not to call for desire but for other pro-responses (Ewing 1939; Anderson 1993; Scanlon 1998). For example, we might think that persons have a particular kind of value that calls fundamentally for responses like *respect* and *preservation*, and indeed that desire isn't the proper response to the special value of persons at all. And more generally, FA theorists might hold that the pro-response called for by a valuable thing can vary according to the kind of thing at issue. For example, as Scanlon (1998) suggests, the pro-response called for by a valuable friendship might differ from that called for by a valuable piece of music. Hence, Scanlon (and others) opt to explain facts about value in terms of normative facts about *valuing*, where valuing can consist variously in a range of pro-responses.

But such FA theories face a challenge. As Rabinowicz and Rønnow-Rasmussen (2004) put it, the challenge is to specify "the common denominator for all the responses in the 'pro'-category", to tell us "what makes them 'pro'" (401). What makes respect and reverence, for example, positive responses (as opposed to negative ones)? And why is protecting something a pro-response to it while pondering is not? A possible and extensionally plausible answer would be to define positive responses as responses that are called for by valuable things (and negative responses as responses that are called for by disvaluable things), but this makes FA theories circular. The challenge is for FA theorists to specify what makes pro-attitudes "pro" without circularity.

An early version of this challenge was pressed by W.D. Ross (1939), as a criticism of Ewing's (1939) FA theory of value *simpliciter*. Ross suggests that Ewing's account of value in terms of fitting positive

emotional responses is circular since the positive emotional responses he appeals to are always accompanied by the thought that the object of the response is valuable or "good". Since positive emotional responses to a thing involve taking the thing to be good, the goodness of a thing can't be non-circularly explained by appeal to (fitting) positive emotional responses. Ewing's (1947) response was to insist that positive emotional responses aren't necessarily accompanied by the thought their object is good. Instead, a positive emotional response need only be accompanied by thoughts about its object's non-evaluative, *good-making* characteristics. Hence, positive emotional responses needn't involve the thought that their objects are good.

The success of Ewing's reply to Ross continues to be a matter of debate (Hurka 2014, ch. 2). In any case, Ewing's reply doesn't offer a positive account of what makes pro-responses "pro"; it merely rejects a possible answer to that question that would make FA theories circular. Another possible response to the challenge is to embrace circularity. According to some, even if we can't adequately characterize pro-responses without reference to the kind of value that FA theories seek to explain, FA theories can nonetheless remain illuminating and informative (see esp. Wiggins 1987 and Tappolet 2011). In short, we can concede the circularity, but claim it's not vicious. This response pairs well with the "no-priority view" discussed in section 2.1, but seems contrary to FA theories of value, insofar as their explanatory component is essential to them.

A third response is to concede the purported circularity of certain FA theories of value *simpliciter*, to reject such theories in light of that circularity, but to embrace FA theories of more specific evaluative properties, which, on the face of it, seem less vulnerable to circularity worries. According to this response, the right response to the present challenge isn't to abandon FA theory *tout court*, but to be more selective about which FA theories we accept. For example, we might reject an FA theory of value *simpliciter* on grounds of circularity, but accept FA theories of being delicious, trustworthy, admirable, despicable, lovable, delightful, and so on. Of course, proponents of even this more limited approach may still ultimately need to characterize the responses in terms of which they seek to explain these evaluative properties without referring to those properties, but, according to some, the prospects for doing so seem at least comparatively promising (Gibbard 1990; D'Arms and Jacobson 2000b; Rabinowicz and Rønnow-Rasmussen 2004; Rosen 2015a).

A fourth and final response, of course, would be to try to tackle the circularity challenge head on, by offering a plausible account of what makes pro-responses "pro" that doesn't refer to the notion of value *simpliciter* that certain FA theories seek to explain. Perhaps, for example, what makes pro-responses "pro" is some common descriptive motivational element (Rabinowicz and Rønnow-Rasmussen 2004). Or perhaps there's some common normative element to all pro-responses that FA theorists could non-circularly appeal to. At the time of writing, however, this issue is underexplored, and the task of developing this kind of response remains outstanding.

Bibliography

- Anderson, Elizabeth, 1993, *Value in Ethics and Economics,* Cambridge, MA: Harvard University Press.
- Brandt, Richard, 1946, "Moral Valuation", *Ethics*, 56: 106-21.
- ——, 1979, *A Theory of the Good and the Right*, Prometheus Books.
- Brentano, Franz, 1889/2009, *The Origin of Our Knowledge of Right and Wrong*, Oxford: Routledge.
- Broad, C.D., 1930, *Five Types of Ethical Theory*, London: Routledge & Kegan Paul.
- Bykvist, Krister, 2009, "No Good Fit: Why the Fitting Attitude Analysis of Value Fails", *Mind*, 118: 1-30.
- Chappell, Richard, 2012, "Fittingness: The Sole Normative Primitive", *Philosophical Quarterly*, 62: 684-704.
- Chisholm, Roderick, 1986, *Brentano and Intrinsic Value*, Cambridge University Press.

- Crisp, Roger, 2000, "Value … And What Follows by Joel Kupperman", *Philosophy*, 75: 452-462.
- Dancy, Jonathan, 2000, "Should We Pass the Buck?", Royal Institute of Philosophy Supplement, 47: 159-73.
- Danielsson, Sven, and Jonas Olson, 2007, "Brentano and the Buck-Passers", *Mind*, 116: 511-22.
- D'Arms, Justin, and Daniel Jacobson, 2000a, "Sentiment and Value", *Ethics*, 110: 722-48.
- ——, 2000b, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophy and Phenomenological Research*, 61: 65-90.
- Dreier, James, 1993, "Structures of Normative Theories", *The Monist*, 76: 22-40.
- Ewing, A.C., 1939, "A Suggested Non-Naturalistic Analysis of Good", *Mind*, 48: 1-22.
- ——, 1947, *The Definition of Good*, London: Routledge & Kegan Paul.
- ——, 1959, *Second Thoughts in Moral Philosophy,* London: Routledge & Kegan Paul.
- Aaron, Elliott, 2017, "Reasons, Dispositions, and Value", *Philosophers' Imprint*, 17.
- Fine, Kit, 1994, "Essence and Modality", *Philosophical Perspectives*, 8: 1-16.
- Foot, Philippa, 1985, "Utilitarianism and the Virtues", *Mind*, 94: 196-204.
- Geach, Peter, 1965, "Good and Evil", *Analysis*, 17: 33-42.
- Gibbard, Alan, 1990, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA: Harvard University Press.
- Howard, Christopher, 2018, "Fittingness", *Philosophy Compass*, 13: e12542.
- ——, 2019, "The Fundamentality of Fit", *Oxford Studies in Metaethics*, 14: 216-36.
- Hurka, Tom, 2001, *Virtue, Vice, and Value*, New York: Oxford University Press.
- ——, 2014, *British Ethical Theorists from Sidgwick to Ewing*, Oxford: Oxford University Press.
- Hurley, Paul, 2017, "Why Consequentialism's 'Compelling Idea' is Not", *Social Theory and Practice*, 43: 29-54.
- Jackson, Frank, 1998, *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Oxford University Press.
- Johnson King, Zoe, 2019, "We Can Have Our Buck and Pass It, Too", *Oxford Studies in Metaethics*, 14: 167-89.
- Kant, Immanuel, 1785/1996, *Groundwork of the Metaphysics of Morals*, trans. and ed. by Mary Gregor, Cambridge: Cambridge University Press.
- Kelly, Tom, 2002, "The Rationality of Belief and Some Other Propositional Attitudes", *Philosophical Studies*, 110: 163-96.
- Kjellmer, Göran, 1986, "Legible but Not Readable: On the Semantics of English Adjectives in –ble", *Studia Neophilologica*, 58, 11-38.
- Leary, Stephanie, 2020, "Grounding the Domain of Reasons", *Australasian Journal of Philosophy*, 98: 137-52.
- Lemos, Noah M., 1994, *Intrinsic Value: Concept and Warrant*, Cambridge: Cambridge University Press.
- Lemos, Noah, 2011, "Intrinsic Value and the Partiality Problem", *Philosophy and Phenomenological Research*, 82: 697-716.
- Lewis, David, 1989, "Dispositional Theories of Value, *Aristotelian Society Supplementary Volume*, 63, 113-37.
- Louise, Jennie, 2004, "Relativity of Value and the Consequentialist Umbrella", *Philosophical Quarterly*, 54: 518-36.
- Maguire, Barry, 2016, "The Value-Based Theory of Reasons", *Ergo: An Open Access Journal of Philosophy*, 3: 1-30.
- McDowell, John, 1998, *Mind, Value, and Reality*, Cambridge, MA: Harvard University Press.
- McHugh, Conor, and Jonathan Way 2016, "Fittingness First", *Ethics*, 126: 575-606.
- ——, forthcoming, *Getting Things Right: Reasons, Value, and Fittingness*, Oxford: Oxford University Press.
- Moore, G.E., 1903, *Principia Ethica*, Cambridge: Cambridge University Press.

- Nye, Howard, David Plunkett, and John Ku, 2015, "Non-Consequentialism Demystified", *Philosophers' Imprint*, 15: 1-28.
- Olson, Jonas, 2009, "Fitting Attitude Analyses of Value and the Partiality Challenge", *Ethical Theory and Moral Practice*, 12: 365-378.
- Orsi, Francesco, 2013a, "Fitting Attitudes and Solitary Goods", *Mind*, 122: 687-98.
- ⸺, 2013b, "What's Wrong with Moorean Buck-Passing?", *Philosophical Studies*, 164: 727-46.
- Parfit, Derek, 2011, *On What Matters*, Oxford: Oxford University Press.
- Piller, Christian, 2006, "Content-Related and Attitude-Related Reasons for Preferences", *Royal Institute of Philosophy Supplement,* 59: 155-82.
- Portmore, Douglas, 2005, "Combining Teleological Ethics with Evaluator Relativism: A Promising Result", *Pacific Philosophical Quarterly*, 86: 95-113.
- ⸺, 2011, *Commonsense Consequentialism: Wherein Morality Meets Rationality*, New York: Oxford University Press.
- Rabinowicz, Wlodek and Toni Rønnow-Rasmussen, 2004, "The Strike of the Demon: On Fitting Pro-Attitudes and Value", *Ethics*, 114: 391-423.
- Reisner, Andrew, 2009, "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem", *Philosophical Studies,* 145: 257-272.
- ⸺, 2015, "Fittingness, Value, and Trans-World Attitudes", *Philosophical Quarterly*, 65: 464-85.
- Rosen, Gideon, 2015a, "The Alethic Conception of Moral Responsibility", in R. Clarke, M. McKenna, and A. Smith (eds.), *The Nature of Moral Responsibility: New Essays*, New York: Oxford University Press, pp. 65-88.
- ⸺, 2015b, "Real Definition", *Analytic Philosophy*, 56: 189-209.
- Ross, W.D., 1939, *Foundations of Ethics*, Oxford: Clarendon Press.
- Rowland, Richard, 2016, "In Defence of Good *Simpliciter*", *Philosophical Studies*, 173: 1371-91.
- ⸺, 2019, *The Normative and the Evaluative*: *The Buck-Passing Account of Value*, Oxford: Oxford University Press.
- Scanlon, T.M., 1998, *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
- Schroeder, Mark, 2009, "Buck-passers' Negative Thesis", *Philosophical Explorations*, 12: 341-7.
- ⸺, 2007a, *Slaves of the Passions*, New York: Oxford University Press.
- ⸺, 2007b, "Teleology, Agent-Relative Value, and 'Good'", *Ethics*, 117: 265-95.
- ⸺, 2010, "Value and the Right Kind of Reason", *Oxford Studies in Metaethics*, 5: 25-55.
- ⸺, 2012, "The Ubiquity of State-Given Reasons", *Ethics*, 122: 457-88.
- Shah, Nishi, 2006, "A New Argument for Evidentialism", *Philosophical Quarterly*, 56: 481-498.
- Sharadin, Nathaniel, 2013, "Schroeder on the Wrong Kind of Reasons Problem for Attitudes", *Journal of Ethics and Social Philosophy,* 7: 1-8.
- ⸺, 2015, "Reasons: Wrong and Right", *Pacific Philosophical Quarterly*, 97: 371-99.
- Sidgwick, Henry, 1874/2011, *The Methods of Ethics*, Cambridge University Press.
- Skorupski, John, 2010, *The Domain of Reasons*, Oxford: Oxford University Press.
- Smith, Michael, 1994, *The Moral Problem*, Blackwell.
- ⸺, 2003, "Neutral and Relative Value after Moore", *Ethics*, 113: 576-98.
- ⸺, 2009, "Two Kinds of Consequentialism", *Philosophical Issues*, 19: 257-72.
- Stratton-Lake, Philip, 2005, "How to Deal with Evil Demons: comment on Rabinowicz and Rønnow-Rasmussen", *Ethics*, 115: 788-98.
- Stratton-Lake, Philip and Brad Hooker, 2006, "Scanlon versus Moore on Goodness", in T. Horgan and M. Timmons (eds.), *Metaethics after Moore*, Oxford: Oxford University Press, pp. 149-68.
- Suikkanen, Jussi, 2009, "Buck-Passing Accounts of Value", *Philosophy Compass*, 4: 769-779.
- Svavarsdóttir, Sigrun, 2014, "Having Value and Being Worth Valuing", *Journal of Philosophy,* 111: 84-109.
- Tappolet, Christine, 2011, "Values and Emotions: Neo-Sentimentalism's Prospects", in C. Bagnoli (ed.), *Morality and the Emotions*, New York: Oxford University Press, pp. 117-134.

- ––, 2016, *Emotions, Values, and Agency*, Oxford: Oxford University Press.
- Thomson, Judith Jarvis, 1993, "Goodness and Utilitarianism", *Proceedings and Addresses of the American Philosophical Association*, 67: 145-59.
- ––, 1997, "The Right and the Good", *Journal of Philosophy*, 94: 273-98.
- Way, Jonathan, 2012, "Transmission and the Wrong Kind of Reason", *Ethics*, 122: 489-515.
- ––, 2013, "Value and Reasons to Favour", *Oxford Studies in Metaethics*, 8: 27-49.
- ––, 2020, "A Puzzle about Enkratic Reasoning", *Philosophical Studies*, https://doi.org/10.1007/s11098-020-01575-z
- Wiggins, David, 1987, "A Sensible Subjectivism?", in *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Basil Blackwell.
- Zimmerman, Michael J., 2001, *The Nature of Intrinsic Value*, Rowman and Littlefield.
- ––, 2011, "Partiality and Intrinsic Value", *Mind*, 120: 447-483.

Academic Tools

[Auto-inserted by SEP staff]

Other Internet Resources

- [Please contact the author with suggestions.]

Related Entries

Franz Brentano | Henry Sidgwick | Intrinsic vs. Extrinsic Value | Value Theory | Value Pluralism

<Christopher Howard>
<chris.howard@mcgill.ca>