The fitting attitudes analysis of value: an explanatory challenge

Kent Hurtig[1]

Abstract This paper is concerned with the implication from value to fittingness. I shall argue that those committed to this implication face a serious explanatory challenge. This argument is not intended as a knock-down argument against FA but it will, I think, show that those who endorse the theory incur a particular explanatory burden: to explain how counterfactual (dis)favouring of actual (dis)value is possible. After making two important preliminary points (about one of the primary motivations behind the theory and what this implies, respectively) I briefly discuss an objection to FA made by Krister Bykvist a few years ago. The point of discussing this objection is to enable me to more easily present my own, and I believe stronger, version of that objection. The overall argument takes the form of, simply, a counterexample which can be constructed on the back of (an acceptance) of my two preliminary points. Throughout the paper I try to respond to various objections.

Keywords Value · Fittingness · Normativity · Non-actual evaluation

According to the fitting attitudes (FA) account of value to be valuable is to be a fitting object of a pro-attitude. A consequence of this theory is that

For all x, x is valuable if and only if it is fitting to favour x.[1]

---

[1] Favouring is here understood as a broad term covering various kinds of pro-attitudes (including pursuit and promotion). The theory is usually understood to be primarily a theory of what is sometimes called the various *modes* of value—i.e. ways in which something may be valuable (for its own sake, for the sake of its effects, for the sake of the contribution it makes to a [valuable] whole etc.).

✉ Kent Hurtig
kent.hurtig@stir.ac.uk

1 Law and Philosophy, University of Stirling, Stirling FK9 4LA, UK

A standard objection to the theory questions the implication from right-to-left: there seem to be cases where it would be fitting to favour an object for its own sake even though it lacks value altogether.[2] This paper is not concerned with this problem; rather, it is concerned with the opposite implication of the biconditional above-the one from value to fittingness. I shall argue that anyone who is committed to this implication faces a very difficult explanatory challenge. This argument is not intended as a knock-down argument against FA but it will, I think, show that those who endorse the theory incur a particular explanatory burden: to explain how counterfactual (dis)favouring of actual (dis)value is possible. After making two important preliminary points I briefly discuss Krister Bykvist's 'solitary goods' objection to FA. The point of discussing this objection is to enable me to more easily present my own, and I believe stronger, version of that objection. The overall argument takes the form of, simply, a counterexample which can be constructed on the back of (an acceptance) of my two preliminary points. Throughout the paper I try to anticipate possible objections and respond to them.

So, to the first preliminary point, which concerns a prominent motivation for the FA analysis. In their seminal discussion of the FA analysis of value, Wlodek Rabinowicz and Toni Rønnow-Rasmussen say that a large part of the attraction behind FA is that it

> establishes a connection between the axiological and the deontic notions: value on this approach is explicated in terms of the stance that *should* be taken toward the object. That it is fitting to have a certain attitude, that there are *reasons* to have it, or that the attitude in question is appropriate or called for, are different ways to express this deontic claim. Consequently, an important advantage of [FA] is that it removes the air of mystery from the *normative* 'compellingness' of values. There is nothing strange in the prescriptive implications of value ascriptions if value is explicated in deontic terms. (My italics.)[3]

---

[2] The standard example in the literature is that a demon will harm you unless you favour something which is stipulated as being completely devoid of value (like a saucer of mud). In this case it seems to be fitting to favour the saucer of mud, yet the saucer of mud is, *ex hypothesi*, not valuable. The problem is that the fittingness of favouring the saucer of mud in this case is of the wrong kind as far as FA is concerned, and the challenge is to provide a principled way of distinguishing the right kind of fittingness from the wrong kind. This has turned out to be surprisingly difficult to do. The problem has become known as the 'wrong kind of reason' problem because it emerged in the wake of T. M. Scanlon's endorsement of what he calls 'the Buck-passing account' of value. According to Scanlon: 'To call something valuable is to say that it has other properties that provide reasons for behaving in certain ways with regard to it.' (1998, p. 95). Some proponents of the FA analysis of value argue that it's a distinct advantage of the FA theory that it is not vulnerable to the wrong kind of reason problem (see for instance McHugh and Way 2016). Although I think this is not obviously so, because there may be good reason to believe that the Buck-passing account properly understood is extensionally equivalent to the FA analysis. In any case, there are obvious affinities between the two theories.

[3] Rabinowicz and Rønnow-Rasmussen (2004), pp. 391–392

It's not clear that 'fittingness', 'appropriate' and 'called for' are *deontic* terms (like 'obligation' and 'duty' etc.) rather than *normative* terms.[4] I suspect Rabinowicz and Rønnow-Rasmussen actually mean 'normative' rather than 'deontic'—after all, an important advantage of the theory, as they put it, is that it explains the *normative* compellingness of value (note also the explicitly normative terms I italicised in the passage above). Thus understood, the FA analysis says that value should be understood in normative terms.

The second important preliminary point is that it is widely, though not universally, accepted that *normativity implies can*. The most famous example of this is of course the *Ought Implies Can* principle, but similar principles seem to hold for other normative terms as well. It plausibly holds for reasons, and, importantly, for fittingness as well.[5] The main thought behind all such principles is very likely the idea that normative facts (or 'considerations') must be able to be action-guiding in some relatively straightforward sense. A normative fact can in turn be action guiding only if it is possible in some relevant sense to do (or believe, or feel etc.) that which the normative fact bids.[6] I will not be able to argue for these claims here, but on the assumption that normativity does indeed imply can, the FA analysis of value entails that

For all x, x is valuable only if x can be favoured.

This implication in turn immediately invites questions about the modality expressed by 'can' in the consequent. What kind of modality are we talking about here? Fortunately for my purposes (as will hopefully become clear later) I need not provide an answer to this question. We can simply assume that the kind of modality expressed by the 'can' in the consequent of the conditional above is at least as strong as the modality expressed by the 'can' in 'ought implies can'—whatever that turns out to be.

---

[4] Some philosophers use 'normative' as a broad term for all those terms that moral philosophers have traditionally been interested in: *good, bad, duty, obligation, requirement, reason(s), fittingness, ought* etc. Other philosophers use 'evaluative' in the same sense (as is usually the case when the fact/value distinction is invoked). Using this broad sense of either 'normative' or 'evaluative' is of course fine for certain philosophical purposes – e.g. when investigating whether the normative/evaluative (thus understood) is reducible to the non-normative. However, we must recognise that there is also a narrower sense of both 'normative' and 'evaluative'. When used in this narrower sense, claims like 'the evaluative is reducible to the normative' and 'the deontic is reducible to the evaluative' are not trivially true. Sorting the italicised terms above into these narrower types is no easy task and there is bound to be disagreement here. Nonetheless, it seems to me that the narrower sense of 'evaluative' encompasses at least 'good', 'bad', 'better', and 'worse'; the narrower sense of 'normative' encompasses at least 'reason(s)', 'ought', and 'fitting'; and the deontic encompasses at least 'duty' and '(moral) obligation'. Thanks to an anonymous referee for pressing me on this.

[5] McHugh and Way (2016) argue that 'fittingness [is] the basic normative property from which the rest of the normative and evaluative domain is constructed' (p. 576, my italics). For arguments that there is reason for someone to F only if that person can F, see Streumer (forthcoming) and Vranas and Peter (2007).

[6] Schroeder's (2007) 'surprise party' example (Ch. 1), and others like it, complicate matters somewhat. For an argument that a version of the 'reasons imply can' principle is not threatened by such examples, see Way and Whiting (2016).

Preliminaries over, we can now focus on the problematic aspect of the FA analysis of value that I wish to highlight. To help us do so, let us start by considering a slightly modified version of Krister Bykvist's 'solitary goods' example:

> There are happy egrets but no past, present, or future agents who are capable of having favouring attitudes.[7]

Assuming, plausibly, that it's a good state of affairs that there are happy egrets, Bykvist asks what favouring attitude it is possible to fittingly have toward this state of affairs. He discusses and dismisses a range of candidate attitudes (intentional pursuit, desire [as the disposition to intentionally bring about] and truth-or belief-entailing propositional pleasure) on the grounds that the having of these attitudes as a response to the happy egrets is either impossible or irrational, and as such these attitudes cannot be *fittingly* held.[8]

An immediate reaction to Bykvist's example, which he himself recognises, is to suggest that the past, present, and future absence of agents is not really relevant to the value of the happy egrets. What is valuable is either their happiness or the state of affairs [There are happy egrets]. The absence of agents is neither here nor there. The state of affairs A: [There are happy egrets] is positively valuable, whereas the state of affairs B: [There are no past, present, or future agents] is of neutral value. If the state of affairs C (= A+B) is valuable, then surely that is so only because C derives its value from A. So a possible response to the example would be to distinguish between non-derived (or basic) and derived final value and claim that proponents of FA should say only non-derived value merits pro-attitudes. Thus understood, the FA account of value would say

> To be non-derivatively (or basically) valuable is to be a fitting object of a pro-attitude

This theory is not vulnerable to the 'solitary goods' objection since it is not impossible or irrational to, say, take pleasure in there being happy egrets. However, as Bykvist points out, this would be

> a radical revision of the original proposal since the things we often evaluate are wider states of affairs, such as situations, outcomes, and lives, and perhaps even whole possible worlds, none of which [are likely to be non-derivatively valuable]. It does not seem plausible to say that good things of this kind do not

---

7 Bykvist (2009), p. 5. Bykvist assumes, plausibly, that states of affairs are value-bearers, but he is open to the possibility of there being other kinds of value-bearers as well.

8 Very briefly, it is not fitting, because impossible, to intentionally bring about, or to be disposed to intentionally bring about, a state of affairs in which nothing has been, is, or will be intentionally brought about. Nor is it fitting to take pleasure in a solitary good if taking such pleasure entails either that a: the object of one's pleasure obtains – as this could not accommodate the value of the state of affairs [There are happy egrets but no one takes (propositional) pleasure in anything]; or b: one believes that the object of one's pleasure obtains – as this could not accommodate the value of the state of affairs [There are happy egrets but there are no believers]. Throughout his discussion, Bykvist also seems to assume something very similar to what I have called *normativity implies can*.

merit our favouring. Much of the attractiveness of the FA-account is therefore lost by limiting its application in this way.[9]

We should not, therefore, accept this restriction on the FA analysis and demand that any plausible version of it should allow that derivative value does not preclude final value.

A plausible response to Bykvist's example is provided by Francesco Orsi. Labelling the solitary goods case 'q' he argues that

> All we need to do in this case is to simply imagine or contemplate *q* as a mere possibility: surely it is then fitting to take what we may call contemplative pleasure in *q*. […] But [having this attitude] does not require believing that [the happy egrets] actually exist (or their actually existing). …the fittingness of the merely contemplative attitudes…does not require that the contemplation and the contemplated state of affairs possibly belong to the same world. […] Since there is no modal requirement that the (dis)favouring world and the (dis)favoured world possibly coincide, the [goodness or] badness of a [state of affairs in a non-actual world] can be accounted for by the fittingness of responding with contemplative [pleasure or] displeasure from a different world.[10]

The idea is that responding with merely contemplative pleasure toward a non-obtaining state of affairs does not 'insert' the contemplator into the contemplated state of affairs and as such it is not irrational or impossible to fittingly favour solitary goods.

This may be exactly what defenders of FA should say about non-obtaining states of affairs (like the solitary goods cases that Bykvist discusses), but it is far from clear what they should say about the value or disvalue of *actually obtaining* states of affairs which are such that it is impossible for anyone in the actual world to have a (dis)favouring attitude toward them. Consider the following conjecture:

> (C) *p* is a significant true proposition, and no one in the actual world at any time has any attitude toward *p*

It is overwhelmingly plausible that there is some *p* for which (C) is true. Since *p* is a significant true proposition—a proposition it would be good to know—corresponding to (C) will be the *actually* obtaining *bad* state of affairs, *S*

> [*p*, and no one in the actual world at any time has any attitude toward *p*]

S's *actually* obtaining is surely bad for its own sake. At the very least, it is *coherent* to think that *S*'s obtaining is bad for its own sake. Furthermore, and very importantly, it is coherent to think that *S's* obtaining *in the actual world* is bad for its own sake. How can this be accounted for by FA? None of Bykvist's candidate attitudes will work (for the reasons he points out), and nor will Orsi's 'contemplative attitudes' suggestion work since it is impossible to have any kind

---

of attitude—including contemplative displeasure and the like—toward S. S, and the proposition that corresponds to it is, to paraphrase Frederic Schick, not properly entertainable.[11]

No past, present, or future person in the actual world can have a negative attitude toward S since p is an essential component of S and as stipulated by the example no one in the actual world has ever had, or will ever have, any attitude toward p. I am of course assuming here that one cannot have an attitude toward a particular state of affairs without having *some* attitude (belief?) toward at least salient essential components of that state of affairs—components like p. Without such an attitude, why think that someone could have a negative (or indeed any) attitude toward *that* state of affairs? It's very plausible, or at least coherent, therefore that S obtains, and its obtaining is bad for its own sake.[12]

Perhaps the FA theorist could respond as follows: In order to contemplate the solitary good of the happy egrets (again calling this g) we don't need to single out any one particular (non-actual) world at which g obtains; we need only entertain the proposition that there is *some* world at which g obtains. Now consider some *actual* solitary good, g*. By hypothesis, no actual person can identify, and so no actual person can contemplate, g*. But why can't a non-actual person do so? If contemplating g doesn't require singling out some particular world at which g obtains, why should contemplating g* (or e, or any other actual solitary good or evil) require singling out some particular world?[13]

This reply misses its target. Although it is true that contemplating g does not require singling out some particular world at which g obtains, this is so only because Bykvist's example is not world-specific. As Orsi points out, all we need to do in such a case is to simply imagine or contemplate q as a mere possibility. But this manoeuvre is not available in my example. It is true that a non-actual evaluator can disfavour the state of affairs that corresponds to the general proposition 'There are propositions such that it would be good if they were believed but which at no time will be believed'. This has the same form as Bykvist's original example about the happy egrets, but this 'general' kind of state of affairs is not the one we're in the business of evaluating. What we are interested in is the badness of a state of affairs obtaining in a *particular* world—our world, the actual world. It's very plausible that, for any world in which S obtains, it is bad for its own sake that S obtains in that world, just like it's overwhelmingly plausible that the state of affairs [There are happy egrets but no past, present, or future agents] is a good state of affairs in every world in which it obtains. But since this is so, [S obtains in @]—where @

---

[11] See his 'The logic of Ambiguity', p. 99ff.

[12] In an earlier draft of this paper C was stated in terms no one having a *belief* about p, and I claimed that it was hard to see how anyone could have an attitude toward S without having some belief-like attitude toward p, but as an anonymous referee has pointed out, no one's having a belief about p is still compatible with someone's having some *belief-like* attitude – other than belief – toward p. Having such an attitude toward p would then enable an agent in the actual world to disfavour S. However, by substituting 'attitude' for 'belief' this possibility is, as far as I can see, foreclosed. I am very grateful to the anonymous referee for pressing me on this point.

[13] I'd like to thank an anonymous referee for suggesting this response on behalf of FA.

designates the actual world—is bad for its own sake, just like [*S* obtains in *w*] is bad for its own sake, and [*S* obtains in *w*'] is bad for its own sake, and so on. Even if the disvalue of [*S* obtains in @] is somehow derived from the badness of the state of affairs that corresponds to the general proposition 'There are propositions such that it would be good if they were believed but which at no point will be believed', [*S* obtains in @] is still bad for its own sake; after all, and as we saw earlier, denying this would incur a commitment to the claim that it is impossible for final (dis)value to be derived (dis)value. This is a commitment we should avoid.

Let's remind ourselves that proponents of the FA analysis are not committed to the claim that the evaluating and the evaluated worlds need to coincide. The introduction of the distinction between the evaluating and evaluated world as a response to the problem of solitary goods is plausible, at least in part, because in that case (the happy egrets) it is obvious that the evaluated world is a non-actual world. In the case I have presented, however, the evaluated world is the actual world, but the evaluating world *cannot* be the actual world. So the evaluating—if there is to be any at all—*has* to take place from a non-actual world.

The problem for the FA theorist is to explain how the counterfactual evaluator can single out events in *our* world (which he would have to be able to do in order to fittingly disfavour the obtaining of *S* in the actual world)? What would make the counterfactual evaluation latch on to the disvalue of [*S* obtains in @], as opposed to the disvalue of *S*'s obtaining in any other world in which is does obtain? Paraphrasing Brogaard and Salerno (who are discussing the problem of the possibility of counterfactual knowledge of actual unknown truth), we can put the challenge as follows:

> …if there is such non-actual evaluation, there is non-actual evaluating of an actual situation. So the non-actual evaluator somehow manages to single out an actually obtaining state of affairs. But how is it possible for a non-actual evaluator to single out states of affairs that obtain in this the actual world? Since there is no causal link between the actual world [@] and the relevant non-actual world [w], it is unclear how non-actual evaluation in w can be uniquely about [*S*'s obtaining in @]. Therefore, it is unclear how there could be non-actual evaluation of what is actually disvaluable.[14]

So the challenge is the following: Since it is impossible for actual evaluators to have the requisite attitude toward [*S* obtains in @]—because of the nature of *S*—and counterfactual evaluators have no way of latching on to the disvalue of [*S* obtains in @]—because they have no way of latching onto @—how are we to understand the *normative* implications of value that FA says it has?

Of course, this is a problem only if *S* is genuinely a bad state of affairs, but perhaps this can be denied? An anonymous referee has pointed out that a defender of FA can claim that although it would be good for us to have some relevant attitude toward (the significant proposition) *p* it doesn't follow that our not having any such attitude toward *p* is bad. So proponents of the fitting-attitudes account might simply

---

14 Brogaard and Salerno (2013).

deny that S is disvaluable. There are, as far as I can see, at least two ways of responding to this worry.

First, suppose we grant that S is not disvaluable. Since S is clearly not valuable we may reasonably take it that S must be value neutral. According to the standard FA analysis of value, for something to be value neutral is for that thing to be a fitting object of indifference. Of course, being indifferent toward something is very different to simply not having an attitude toward that thing. There are lots of things I have no attitude toward (things I've never heard of, for example) but it would be a mistake to describe me as being indifferent toward those things. I'm not sure how best to characterise indifference—perhaps indifference should be understood in terms of choice dispositions?—but however it is best understood, it strikes me as very plausible that, for reasons already discussed, no one in the actual world can be indifferent toward S and non-actual evaluators can't be indifferent toward S because they can't pick out S. So the problem reappears with neutral value and the fittingness of indifference.

Second, even if this response somehow fails, we need to keep in mind that, as Bykvist puts it, "the FA-analysis is not just an account of absolute values, goodness, badness, and neutrality, it is also an account of betterness: what is better is what it is fitting to prefer or favour more."[15] If it would be better if someone had a relevant attitude toward p this would mean that there is some state of affairs, S^*, such that S^* is better than S. This in turn entails that it is fitting to prefer S^* to S. Now we're back again with yet another version of the original problem: *Who* can have the relevant preference? No one in the actual world can prefer S^* to S since such a preference is possible only if one has some attitude toward S—which no one in the actual world will ever have and no non-actual evaluator can pick out. So it seems that denying that S is disvaluable will not get round the general problem of explaining how it is possible to have the fitting attitude toward S (or to prefer S^* to S).

The price of evaluating value in normative terms (be in terms of fittingness, reasons, or 'ought') is that normative concepts very plausibly have an 'ability condition' attached to their correct application. In essence, *normativity implies can*, and it is for this reason that proponents of the FA incur the explanatory burden of showing *how* it is possible to favour *specific* worlds, situations, or states of affairs. Compare this to, say, a hedonist theory of value according to which being valuable just is being pleasurable. On such a theory there may well be states of affairs obtaining in specific non-actual worlds that are good because they contain pleasure, but unlike the FA theorist, the hedonist does not have to explain how it is possible to single out such worlds and states of affairs since her theory is not cast in normative terms. It is no embarrassment for the hedonist to acknowledge that there are states of affairs that are valuable for their own sake that no one will ever have any beliefs about, and which, subsequently no one will ever have any favouring attitude toward. So unless the FA theorist has some plausible story to tell about how counterfactual evaluation of actual value is to be conceived of one cannot help but thinking that this is a serious, perhaps fatal, lacuna in the theory.

---

[15] Bykvist, p. 10.

References

Brogaard, B., & Salerno, J. (2013). Fitch's paradox of knowability. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2013 Edition) http://plato.stanford.edu/archives/win2013/entries/fitch-paradox/.

Bykvist, K. (2009). No good fit: Why the fitting attitude analysis of value fails. *Mind, 118*(469), 1–30.

McHugh, C., & Way, J. (2016). Fittingness first. *Ethics, 126,* 575–606.

Orsi, F. (2013). Fitting attitudes and solitary goods. *Mind, 122*(487), 687–689.

Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics*, *114*(April), 391–423.

Scanlon, T. (1998). *What we owe to each other*. Cambridge: Harvard University Press.

Schick, F. (2003). 'The logic of ambiguity' in his *Ambiguity and logic*. Cambridge: Cambridge University Press.

Schroeder, M. (2007). *Slaves of the passions*. Oxford: Oxford University Press.

Streumer, B. (forthcoming). Reasons and ability. In D. Star (Ed.), *The oxford handbook of reasons and normativity*. Oxford: Oxford University Press.

Vranas, P. (2007). I ought, therefore I can. *Philosophical studies, 136*(2), 167–216.

Way, J., & Whiting, D. (2016). Reasons and guidance (or, surprise parties and ice cream). *Analytic Philosophy, 57*(3), 213–245.