

## value and the right kind of reason

Fitting Attitudes accounts of value analogize or equate being good with being desirable, on the premise that ‘desirable’ means not, ‘able to be desired’, as Mill has been accused of mistakenly assuming, but ‘ought to be desired’, or something similar. The appeal of this idea is visible in the critical reaction to Mill, which generally goes along with his equation of ‘good’ with ‘desirable’ and only balks at the second step, and it crosses broad boundaries in terms of philosophers’ other commitments. For example, Fitting Attitudes accounts play a central role both in T.M. Scanlon’s [1998] case *against* teleology, and in Michael Smith [2003], [unpublished] and Doug Portmore’s [2007] cases *for* it. And of course they have a long and distinguished history.

A major stumbling block to Fitting Attitude accounts (henceforth: ‘FA accounts’) of anything, however, is the famous *Wrong Kind of Reasons* problem.<sup>1</sup> The problem is intuitively this: it may sometimes be a good idea for me to desire something bad – for example, if I will get some significant reward for doing so. If the reward is high enough, then it may even be that I *ought* to desire the bad thing. But that doesn’t make the bad thing good. So some things ought to be desired without being good. Conversely, if the reward attaches to *not* desiring something, even though it is good, we can get cases in which something is good, even though it is not the case that I ought to desire it. Either way, ‘good’ cannot, it seems, mean ‘ought to be desired’ after all. This is called the ‘Wrong Kind of Reasons’ problem because the incentive – the reason – for desiring the bad thing appears to be of the ‘wrong kind’ to contribute to whether that thing is good or not. Some have held that this means that the FA account of ‘good’ must be mistaken.

In this paper I offer a new approach to this problem, characterized by its high level of generality.<sup>2</sup> I claim, by considering the full breadth of cases in which Wrong Kind of Reasons-style problems arise, to offer a better diagnosis of what really lies at the root of the Wrong Kind of Reasons problem. My diagnosis has a payoff: as evidence that it is on the right track, I’ll show that

---

<sup>1</sup> See, for example, D’Arms and Jacobson [2000b] and Rabinowicz and Ronnow-Rasmussen [2004].

<sup>2</sup> The approach given here develops that briefly outlined in chapter seven of Schroeder [2007b] in greater detail.

it solves parallel puzzles that arise in giving FA accounts of ‘good for’ and attributive ‘good’, which do not on their face appear to be versions of the Wrong Kind of Reasons problem at all. It also leads to a moral about what it would take for so-called ‘agent-relative value’ to be a kind of value.

The paper has five parts, stating, respectively, (1) the problem, (2) its scope, (3) my diagnosis, (4) the confirmation through application to the two apparently unrelated problems I mentioned, and (5) the application to the debate about agent-relative value.

## I problem

I take both the basic idea of FA accounts and the basic issue behind the Wrong Kind of Reasons problem to be relatively clear, or if not, at least unclear in familiar ways; in this section of the paper I’ll confine myself to remarks on how optimistic we should be about a solution to the problem, and on why we might expect it to take the form of a distinction between the ‘right kind’ of reasons and the ‘wrong kind’.

### *FA and WKR: everyone’s problem*

It is important to start with the recognition that the Wrong Kind of Reasons problem is not a problem specific to proponents of FA accounts of value. This is because even though FA accounts of value – that is, of ‘good’ and its relatives – are deeply controversial, there are many concepts for which FA accounts are not and should not be deeply controversial.

This is exhibited by the critical reaction to Mill. *Maybe*, this critical reaction goes, Mill is right to assume that ‘good’ means ‘desirable’. But ‘desirable’ most certainly does not mean ‘able to be desired’; it means, rather, ‘ought to be desired’ (or something similar). This critical reaction to Mill is built on the highly compelling idea that an FA account *has* to be right for ‘desirable’. And FA accounts are also highly compelling for a variety of related terms: ‘admirable’, ‘enviable’, ‘laudable’, ‘preferable’, ‘contemptible’, and ‘notable’, just to sample a few with the same suffix, and these are joined by cousins like ‘scary’, ‘awesome’, ‘amusing’, ‘blameworthy’, and so on.

For every one of these concepts, it is easy to construct Wrong Kind of Reasons scenarios. Imagine financial rewards for admiring the despicable or lauding the mediocre, or invent an evil demon with plans to punish your family if you are not afraid of Mickey Mouse or if you aren’t amused by *Schindler’s List*. The scenarios needn’t even be unrealistic; as D’Arms and Jacobson note, there are real-world jokes that are funny even though it is wrong to be amused by them, and true-

to-life situations in which by far the smartest course is to admire someone lacking in admirableness, or to withhold envy from the enviable.

Yet Fitting Attitudes accounts of each of these concepts are highly compelling. And their appeal is straightforward. It is that each of these terms is semantically connected to some attitude in an obvious way: desiring, admiring, envying, lauding, preferring, having contempt for, taking note of, fearing, being in awe of, being amused by, and blaming. So each must have *something* to do with its corresponding attitude. So the question is: *what* does each of these things have to do with the corresponding attitude? And the answer is that no non-normative answer seems to suffice. Not only is desirability not a matter of being able to be desired, it is not a matter of being often desired, of being desired by well-educated people, or a matter of being desired in periods of calm reflection. To the extent that any of these alternative non-normative attempts to capture the relation to desire that it takes for something to be desirable succeed, it is that they merely approximate some *normative* condition on desire.<sup>3</sup> But all it takes to have a Fitting Attitudes account of desirability is to think that for something to be desirable is for some normative condition (*FILL-IN-THE-BLANK*) to hold for its being desired: for it to be *FILL-IN-THE-BLANK*edly desired.

All of this means that the Wrong Kind of Reasons problem is everyone's problem – or at least, close enough to give us comfort. If some sort of Fitting Attitudes account *has* to be right about 'admirable', 'enviable', and so on – or even just some non-empty subset of this list – then there *has* to be some concept of *fittingness* – some answer to what goes for '*FILL-IN-THE-BLANK*' – that can do the job of being unaffected by reasons of the 'wrong kind'. And whatever that notion of fittingness is, that should do for FA accounts of 'good' and its relatives, as well. In short, we should be highly optimistic that whatever the Wrong Kind of Reasons problem teaches us, there is at least some solution to be had.

*why explain fittingness in terms of reasons at all?*

Of course, the simplest answer to the Wrong Kind of Reasons problem is that there *just is* a notion of *fittingness* – the one that makes the FA accounts of 'admirable', 'enviable' and so on true – and that one of the important facts governing it, is that unlike 'ought, all things considered', it is simply unaffected by the kinds of prudential and moral considerations that go into Wrong Kind of

---

<sup>3</sup> Compare Enoch [2004].

Reasons scenarios. On the most straightforward version of this view, fittingness simply doesn't have to do with reasons at all, so it is equally unaffected by reasons of the 'right kind', as well as of the 'wrong kind'. It's just not the kind of normative concept to be affected or determined by reasons. Call this the *sui generis* solution to the Wrong Kind of Reasons problem.

If the *sui generis* solution to the Wrong Kind of Reasons problem is right, then we can rest easy. Though 'ought, all things considered' may be the kind of normative relation to be affected by our reasons, on this view there are simply other normative relations which are not so affected, and the concept of what is *fitting* just happens to be among them; no more needs to be said. If things can really be this easy, then the Wrong Kind of Reasons problem is, as Rabinowicz and Ronnow-Rasmussen report one of the editors of *Ethics* as suggesting, simply an example of 'philosophical bait and switch'; there is a problem for people who think that 'desirable' means 'ought to be desired', but no problem at all for the view that it means 'fittingly desired'.

Now, there is something right about this accusation – as I just argued, we should be highly confident that there *is* a notion of the *fitting* that is immune from the problem, precisely because we should be highly confident that some FA account of 'admirable', 'enviable', and their relatives is correct. But I also think that there is more to the problem than this, and that in particular, an adequate answer needs to be able to distinguish reasons of the right kind from reasons of the wrong kind. I think that, because I think that on an adequate view, 'fitting' *is* going to need to be understood in terms of reasons, and so reasons of the 'wrong kind' are going to need to be ruled out.

So why think that the notion of the 'fitting' is going to need to be understood in terms of reasons? The answer is: for the very same reasons that we should think that 'ought, all things considered' needs to be understood in terms of reasons. The idea of understanding 'ought, all things considered' in terms of reasons comes down to us from Ross [1930], who noted that a variety of different but competing factors play into determining what we ought to do (all things considered). If the only relevant factor is that an action would be a lie, then we ought not to do it. But if that lie would also happen to save a friend's life, then we ought to do it. Why? Because the reason to do it outweighs the reason not to do it. We should understand 'ought, all things considered' claims to be claims about the balance of reasons, on this view, precisely because that best explains why their truth is affected in the way that it is by the balance of the reasons.

The same thing goes for fittingness. If Mary is more steadfast than Claire, then other things being equal – if that is the only relevant consideration in play – she is also more admirable –

that is, it is fitting to admire her more. But if Claire is far more accomplished than Mary, then it may be that Claire is more admirable than Mary – that is, it may be that it is fitting to admire Claire more. The fact that Mary is more steadfast than Claire is a reason to admire her more, and the fact that Claire is far more accomplished is a reason to admire her more. Which one it is fitting to admire more is affected by each of these reasons, and by their relative weights. Finally, we should understand claims of fittingness to be claims about the balance of (the right kind of) reasons, precisely because that best explains why their truth is affected in the way that it is by the balance of (the right kind of) reasons.

If this reasoning is on the right track, and the grounds for understanding fittingness in terms of the balance of the (right kind of) reasons match the grounds for understanding ‘ought, all things considered’ in terms of the balance of the reasons, then a solution to the Wrong Kind of Reasons problem will need to boil down to a way of distinguishing between reasons of the ‘right kind’ and reasons of the ‘wrong kind’ after all. That, at any rate, is what I will look for in parts 2 and 3. But first, a formulaic look at what Fitting Attitudes accounts will look like, if they take my arguments seriously so far.

*some formulaic accounts*

Many Fitting Attitudes accounts are formulated in sloppy or deliberately imprecise ways, so in this section I state what I think they should look like, *modulo* a theory which provides us with a way of distinguishing between the ‘right’ and ‘wrong’ kinds of reasons.

First, for ‘admirable’. Since ‘admirable’ is a gradable adjective, I assume that it is a mistake to try to give an account of it directly. Rather, an FA account of ‘admirable’ should take the form of an FA account of the binary relation, *more admirable than*:

**more admirable** A is more admirable than B just in case the set of all of the (right kind of) reasons to admire A more than B is weightier than the set of all of the (right kind of) reasons to not admire A more than B.

We can then say the same sort of thing about the relationship of ‘admirable’ to ‘more admirable than’ as we would say about the relationship of ‘tall’ to ‘taller than’ or ‘large’ to ‘larger than’:

**admirable** A is admirable just in case A is more admirable than most [in the contextually supplied comparison class].

If this is what an FA account of ‘admirable’ looks like, then the precisely analogous thing will go for ‘good’. Unfortunately, however, the case of ‘good’ is more complicated. Whereas there is only a limited range of different sorts of sentence structures involving ‘admirable’, there is an immense range of kinds of sentences and constructions in which ‘good’ appears. For example, ‘good’ and ‘better’ appear in sentences like all of the following, just to start with twenty-five examples (see chapter six of Ziff [1960] for an impressive but still incomplete catalogue):

Pleasure is good. She is good. She’s a good skier. She’s good at skiing. She’s good in skis. She’s good with children. She’s good for him. She’s good to him. It’s good for him that she came. It’s good for him for her to come. It’s good for him when she comes. Her coming is good for him. She’s good to talk to. She’s good for him to talk to. It’s good for him to talk to her. It’s good for him to talk. Talking is good for him. It’s better for him to talk than to stay silent. It’s better for him to talk than for him to stay silent. It’s good when it rains. It’s good for it to rain. It’s good that it rained. The fact that it rained is good. It’s a good thing that it rained. It’s a good thing for it to rain.

The striking thing about this set of sentences is that no two of them have exactly the same kind of surface structure, taking syntax and argument types into account. So even if there is only a single semantic value for ‘good’ underlying all of these varying kinds of constructions, there is no ‘canonical formulation’ of claims involving ‘good’, to which to attach an FA account (or any other account of value, for that matter).

So what an account of value needs to do instead, is to separate out the question of the nature of value from the question of which of these sentences make claims about value, so understood. We can do this by first distinguishing the property or relation of which we aim to give an FA account, and providing that account, without paying any attention to how it is related to ordinary English. Then, second, once we have an FA account of some relation in hand, we next specify which lexical entry is associated with this property as its semantic value, and what its syntactic properties are. That generates predictions about which English sentences ascribe the property or relation of which we have provided an FA account. And then together with those predictions, the FA account itself allows us to generate predictions about the truth conditions of those sentences.

So here’s how this works in practice. In at least some of these sentences, including ‘it’s good for it to rain’ and ‘it’s good that it rained’, ‘good’ appears to syntactically take one argument, which is given by the following complementizer clause. It is therefore natural to treat it

semantically as a property of propositions (relative to a comparison class), and hence to treat corresponding ‘better’ sentences as ascribing a two-place relation between propositions. The following accounts constitute an FA account of what these relations might be:

- |               |   |
|---------------|---|
| <b>better</b> | For propositions A and B, $A > B$ just in case the set of all of the (right kind of) reasons to prefer A’s being the case to B’s being the case is weightier than the set of all of the (right kind of) reasons to not prefer A’s being the case to B’s being the case. |
| <b>good</b>   | For proposition A and comparison class of propositions c, $g(A, c)$ just in case for most $x$ in c, $A > x$ .   |

Now we can put **better** together with the thesis that there is at least one lexical entry for ‘better’ in English on which it syntactically takes two complementizer clauses for its arguments and has the relation,  $\_ > \_$ , for its semantic value, in order to predict that ‘it is better for him to talk than for him to stay silent’ says that  $A > B$ , where A is the proposition that he talks and B is the proposition that he stays silent. Similarly, we can put **good** together with the thesis that there is at least one lexical entry for ‘good’ in English on which it syntactically takes a single complementizer clause for its overt argument and a contextually salient comparison class for its covert argument, and has the relation,  $g(\_, \_)$  for its semantic value, in order to predict that ‘it’s good for it to rain’ says that  $g(A, c)$ , where A is the proposition that it rains and c is a comparison class provided by the context of utterance. (One of the virtues of this account is that we can delegate the explanation of the factiveness of ‘it’s good that it rained’ to special features of the ‘that’ complementizer, assigning ‘good’ the very same semantic value as in ‘it’s good for it to rain’.)

This is still a rough picture, but it is closer to what a worked-out FA account of ‘good’ should look like, *modulo* a solution to the Wrong Kind of Reasons problem. It offers an FA account of some relation,  $\_ > \_$ , and then pairs that with an assignment of that relation to an English word with a particular syntactic structure, thereby making predictions about which English sentences have readings which involve this relation, and hence about their truth conditions. Together with the appropriate principles from syntax, this account may be sufficient to account for the last seven ‘good’ sentences on my list. But saying more would require both a more careful account of English syntax and a more careful statement of the view just given here, so this much will do for our purposes, here.

## 2 scope

One of the central claims of this paper is to offer a solution to the Wrong Kind of Reasons problem that is distinguished by its generality. So the job of this section is to emphasize the *scope* of the Wrong Kind of Reasons problem. What I'll be arguing, is that competing solutions have been motivated by considering only a very special case, and hence have been led to overgeneralize on some of its idiosyncratic features.

*tying knots, executing endgames, and setting the table*

So what does it take to give rise to a Wrong Kind of Reasons type set of issues? I think it takes very little. My view is that Wrong Kind of Reasons issues arise in any domain which is governed by a standard of correctness.

For example, there are correct and incorrect ways to tie a knot. Laying one piece of rope across another is an incorrect way to tie a knot, as is twisting them one full time clockwise, followed by one and a half twists counter-clockwise. Similarly, there are correct and incorrect ways to execute the Lucena endgame (a winning rook endgame in chess). And there are correct and incorrect ways to set the table for a White House state dinner. Each of these three activities: tying knots, executing the Lucena endgame, and setting the table for a White House state dinner, is thus governed by some standard of correctness.

For each of these activities, however, we can easily imagine incentives to carry out the activity in some other way. An eccentric millionaire might approach you, as you are tying a knot, and offer you thousands of dollars to simply leave one rope lying on top of the other. Terrorists might threaten your family with torture unless you botch your winning rook endgame position. And subversive elements might offer to donate vast sums to your favorite charities (and find you new employment), if only you will put the salad forks next to the soup spoons at the White House state dinner for which you are setting the tables for this evening.

None of these incentives, I hope you will agree with me, affects what is the correct way to tie a knot, what is the correct way to execute the Lucena position, or what is the correct way to set the table for a White House state dinner. So each of these cases has the same structure as a Wrong Kind of Reasons problem – standards of correctness are unaffected by the kinds of incentives that come into play in Wrong Kind of Reasons scenarios.



*narrow, narrower, and narrowest*

The problem with existing approaches to the Wrong Kind of Reasons problem is therefore that they are too narrow in scope. All such existing approaches are specifically tailored to apply only to the difference between the right and wrong kind of reasons *to have certain kinds of mental states*. But since the Wrong Kind of Reasons problem is much more general, we should expect a full solution to respect the full scope of the problem. This is something that existing solutions do not even attempt to do.

The *most* general way that has been offered of distinguishing between right and wrong kinds of reasons is sometimes called the distinction between ‘state-given’ and ‘object-given’ reasons, or alternatively that between ‘attitude-related’ and ‘content-related’ reasons.<sup>4</sup> It is built into the very framework of this distinction that it can only distinguish between the right and wrong kind of reasons to be in mental states (and only a certain kind of mental state). At maximum generality, it can only apply to mental states which have contents or ‘objects’. Now, Rabinowicz and Ronnow-Rasmussen [2004] have questioned whether the distinction between content-related and object-related reasons can even be made out, and succeeding literature has taken up this question. But my worry is much more general. Even if the distinction can be made out, it is of the wrong kind to provide a *general* solution to the Wrong Kind of Reasons problem, because it applies only to mental states, and only to mental states of a certain kind, at that, whereas the Wrong Kind of Reasons problem comes up for activities such as tying knots, executing chess endgames, and setting the table for White House state dinners, none of which are mental states at all.

Now, several other authors have offered solutions to the Wrong Kind of Reasons problem which differ from appeals to the state-given/object-given distinction. But each of these alternative solutions is even narrower in scope than the state-given/object-given distinction. For example, Rabinowicz and Ronnow-Rasmussen explore but don’t quite endorse the idea that we can distinguish between the right and wrong kind of reasons to have attitudes *which represent the reasons for which we hold them*. Their idea is that the properties which are connected to the right kind of reasons ‘appear in the intentional content of favoring as the features on account of which the object is favored’. Whatever the other merits of this idea (Rabinowicz and Ronnow-Rasmussen go on to criticize it, themselves), it applies to an even narrower range of cases than the state-given/object

---

<sup>4</sup> See, for example, Parfit [2001], Piller [2001] and [2006], and Olson [2004] for versions of this idea, and Rabinowicz and Ronnow-Rasmussen for critical discussion.

given distinction, because it applies only to mental states which represent the reasons for which they are held.

Justin D'Arms and Dan Jacobson [2000b] have also offered what looks like a way of distinguishing between the right and wrong kinds of reasons, which I'll follow Gideon Rosen's [unpublished] usage in calling the *alethic* account. The alethic account works by discriminating between reasons which bear on the *truth* of the content of the attitude from those which do not. This account yields intuitive verdicts in many cases that are quite close to the state-given/object-given distinction, but it is again narrower. A proponent of the state-given/object-given distinction, for example, may distinguish between right and wrong kinds of reasons to have a certain *intention*, by noting that the right kind of reasons to intend to do A must also be reasons to do A, whereas the wrong kind of reasons to intend to do A are reasons to intend which are not themselves reasons to do A (compare Kavka [1983]). The only way for the alethic account to apply to this case, is if we assume that the intention to do A has some particular content, and that reasons to do A must bear on whether this content is true. Now, someone who adopts a 'guise-of-the-good' theory of intention and a value-based theory of reasons may endorse precisely this view,<sup>5</sup> but those are substantial commitments to take on, simply in order to generalize the alethic account.

Finally, Pamela Hieronymi [2005] has advocated a fourth approach to the Wrong Kind of Reasons problem. Hieronymi's view employs the concept of a *commitment-constituted attitude*, and distinguishes between the right and wrong kind of reasons only for such attitudes. An attitude is commitment-constituted just in case there is some question, such that answering it amounts to forming that attitude. Hieronymi makes a strong case that her view deals well with the cases of belief and intention, but it is especially narrow. There is certainly no question such that answering it amounts to tying a knot, or to setting the table for a White House state dinner. Yet tying knots and setting the table for White House state dinners are governed by standards of correctness that are immune from the Wrong Kind of Reasons all the same.

---

<sup>5</sup> The idea here is simple: if intending to do A is seeing A as good, and reasons to do A are considerations which bear on whether A is good, then reasons to do A will bear on the truth of the content of intending to do A, and hence the alethic account will diagnose all and only reasons to do A as of the 'right kind' as reasons for intending to do A.

*descriptive versus deep explanatory accuracy*

I believe that we should abandon attempts to distinguish between the right and wrong kinds of reasons which apply only to mental states. Standards of correctness arise in many different domains, so mental states are only a special case.

This is not, however, to say that any of the existing proposals about the nature of the wrong kind of reasons are extensionally incorrect. It may be that there is something essentially right about the state-given/object-given distinction, or about the alethic view, or about Hieronymi's commitment-constituted attitude approach. Certainly nothing I have said here shows that they are not descriptively correct, within some domain. What I've been arguing, however, is that even if they are right in their own domains, there has to be something deeper and more general to be said, because the distinction between reasons which affect correctness conditions and reasons which do not is much more widespread than the domain of reasons for mental states. Consequently, I have been arguing, we should expect a more general solution to be had. To the extent that each of these other views are descriptively correct, we might expect a more general solution to the problem to help us to explain why.

### **3 diagnosis**

In this section I'll outline my more general solution to the Wrong Kind of Reasons problem and illustrate how it can be applied to some of the cases that we've considered.

*two kinds of WKR*

The first step toward my solution to the Wrong Kind of Reasons problem is to distinguish between two different kinds of reason that come up in 'wrong kind' scenarios. The paradigmatic kind of 'wrong kind' of reason comes up in cases in which someone engaged in some activity is confronted with extraordinary incentives to pursue the activity in some incorrect way – for example, under threat of his life from an evil demon, or at the offer of a huge financial reward from the eccentric billionaire. The striking fact about these sorts of examples, is that though they affect what *some* person has reason to believe, or intend, or feel, or do, they don't affect what *anyone* has reason to do in what is otherwise the same situation.

For example, if it is *your* son who is on trial for murder, and *you* would suffer greatly if you believed he was guilty, that may give *you* a reason not to believe that he is guilty, but it doesn't

given anyone *else* a reason to believe that he is not. Likewise, when the terrorists threaten to kill your family unless you put the salad forks next to the soup spoons in the place settings for the White House state dinner, that may give you a reason to put the salad forks next to the soup spoons, but no one else who is setting the table for a White House state dinner will feel any such pressure. The case of wrong kinds of reason for intentions is harder, but I think the same verdict goes: when the eccentric billionaire offers you a million dollars (to be received at midnight tonight) for intending to drink the toxin tomorrow, that gives you an extra reason to form that intention *now*, but tomorrow morning it will not be a reason for you to have the intention. So it is a reason only for you to have the intention only at certain *times*, and not at any time at which that intention is relevant for deciding what to do.

All of these wrong kinds of reason are *idiosyncratic*, in the sense that they apply to only some of the people who are engaged in the relevant activity – whether that activity is believing, or intending, or setting the table for a White House state dinner. In contrast, some examples of wrong kinds of reasons are not idiosyncratic in this way. The best examples are those of *moral* reasons, as in the cases raised by D’Arms and Jacobson [2000a]. There may be a moral reason not to laugh at or be amused by some joke, for example (D’Arms and Jacobson don’t give any particular examples, but you know what they are like), but that doesn’t affect how amusing the joke is. Since the fundamental moral reasons are categorical, they apply to any agent, no matter what they are like, and no matter what they are doing. Whereas the idiosyncratic reasons apply to *fewer* people than are engaged in the relevant activity, moral reasons apply to *more*.

*the solution: reasons special to an activity*

My solution to the Wrong Kind of Reasons problem is to put these two observations together with an answer to the question, ‘whose reasons?’ When someone tells us that they believe that the desirable is what ought to be desired, we should be asking, ‘desired by whom? Who ought to desire it?’ Similarly, when someone tells us that the desirable is what there is reason to desire, we should be asking, ‘reason for whom to desire it?’ The observation that many Wrong Kind of Reasons examples are reasons that only certain people have and not others, strongly suggests that the answer to this question may have something to do with the right answer to the Wrong Kind of Reasons problem. So that is my idea. Here is what it looks like as a slogan:

**RKR** The right kind of reasons with respect to any activity, A, are the reasons that are shared by necessarily anyone who is engaging in A, and just because they are engaging in A.

According to this account, the distinction between the right and wrong kind of reasons is relative to an ‘activity’. This is because the point of the distinction between the ‘right’ and ‘wrong’ kinds of reasons, is that only the ‘right’ kind contribute to standards of correctness, and standards of correctness are relative to activities. Placing one’s rook in the third rank is an incorrect way to execute the Lucena rook endgame position,<sup>6</sup> but it is the correct way to throw a match – so the very same action can be the correct way of doing one thing but an incorrect way of doing another thing. So since correctness is relative to activities, so is the ‘right kind’ of reasons. Activities, so understood, include tying knots, executing the Lucena endgame, and setting the table for White House state dinners, but I’ll also understand believing, intending, preferring, and admiring as ‘activities’, in the broad sense appropriate to this account.

The requirement that the right kind of reasons be shared by anyone who is engaged in an activity rules out idiosyncratic reasons as being of the wrong kind (the ‘necessarily’ requirement ensures that reasons don’t fail to be idiosyncratic only because the terrorists threaten the only person who ever sets the table for White House state dinners, and similarly for similar cases). The requirement that they be reasons that are shared *just because* those people are engaged in the activity rules out reasons that they would have had anyway – such as the moral reasons in D’Arms and Jacobson’s examples.

These restrictions on whose reasons the right kind must be are not *ad hoc*, because they fall out of a natural picture. Whenever an activity has the feature that it gives rise to a shared set of reasons, there will be an interesting question as to what is required by those reasons, not taking others into account. Such questions are questions about the standard of correctness for the activity. Not all activities give rise to such shared reasons, and hence not all activities are governed by standards of correctness. But many activities do give rise to shared sets of reasons, and plausibly the examples that we have been considering are among them.

This means that we can use our account of the right kind of reasons in order to give an account of *correctness*:

---

<sup>6</sup> The key Lucena position is the concept of ‘building a bridge’ so that one’s king can safely get out of the way of one’s advanced pawn and into the safety of being protected by one’s rook. It is too far to ‘build a bridge’ to a rook on the third rank, because it leaves the advanced pawn up for capture, leading to an easy draw. See, for example, [http://en.wikipedia.org/wiki/Lucena\\_position](http://en.wikipedia.org/wiki/Lucena_position).

**correct**        B is the correct way to A just in case the set of all of the reasons to do B that are of the right kind with respect to activity A is weightier than the set of all of the reasons to not do B that are of the right kind with respect to activity A.

With the account of correctness in hand, we can work backwards, and frame our FA accounts directly in terms of correctness, rather than in terms of the right kind of reasons. (I won't do that here.) Correctness is the right way to fill in the blank for *FILL-IN-THE-BLANK*.

*applications: tying knots and admiration*

By now you should be wondering whether this approach to the Wrong Kind of Reasons problem generates plausible results. Is it really the case that tying knots and admiring are activities that give rise to shared sets of reasons? I think the answer is 'yes'. Moreover, I think that they give rise to shared sets of reasons in different ways, and that this illustrates the flexibility of the solution that I am offering.

Take the case of tying knots first. Tying knots is an *intentional* activity. So anyone who is engaged in it must have certain aims. If you don't have at least some minimal aim that the pieces of rope (or string, or thread) which you are tying together achieve some kind of minimal cohesion, then you may be fiddling with rope, or pretending to tie a knot, but you are not really tying a knot. To really be engaged in the activity of tying knots, you must be at least trying to some minimal extent. But if that is right, then we can infer some interesting things about the reasons that are shared by people who are tying knots, because they are engaged in the activity of tying knots. They will share any reasons that arise from the aim to get two pieces of rope (or string, or thread) to achieve some kind of cohesion. Laying one rope across the other and then letting go is not a correct way to tie a knot because, as observation confirms, it is strikingly unsuccessful in a very wide range of cases at getting the two pieces of rope to stick to one another. So it foils the aim of tying knots, and hence is something that there is much weightier reason of the right kind *not* to do, than to do.

Now, this sort of explanation requires some relatively rich assumptions about what is involved in participating in the activity of tying knots – that it involves trying, to at least some minimal degree, to do a particular thing. So you might worry that it is possible to do things which look on the outside a lot *like* tying knots, but do not involve trying to do this particular thing. That's fine – activities come cheap, and standards of correctness are relative to activities. It is true

that not all activities give rise to shared sets of reasons, but not all activities are governed by standards of correctness. All that we need, is that the ones that are governed by standards of correctness give rise to shared sets of reasons. Activities come cheap, and only when understood in relatively rich ways are they governed by standards of correctness. Fortunately, when an activity is understood in a relatively rich way, that gives us material that we can use in order to explain the shared set of reasons that the activity gives rise to.

So much for the right kind of reasons to tie a not. Admiring is I think a more complicated case.<sup>7</sup> Admiring Max might be the incorrect way to admire, but the correct way to placate some billionaire or evil demon – so when we talk about the right kind of reasons to admire Max (or not), we are interested in reasons which bear on whether admiring Max is the correct way to *admire*. *Admiring*, therefore, is the activity that we must test, to see if it gives rise to a shared set of reasons of the appropriate kind.

But now, the general formula for the account of the right kind of reasons for tying knots is not going to work. There I gave a general recipe that works for the standard of correctness of any *intentional* activity. But admiring is not an intentional activity – it is a kind of psychological attitude. Or rather, though there is such a thing as an intentional activity of admiring – as in ‘I was just admiring all of the work you’ve put into your garden’ – there is *also* an attitude of admiring that is not itself an intentional activity – and it is this one that is needed, here. Since admiring is not an intentional activity, I doubt that there are any specific desires that must be shared by anyone who is engaged in it – that is, by anyone who is in the business of admiring. But I still think that there are going to be some important reasons that are shared by anyone who is engaged in admiring. Here I’ll try out two different ideas that would get us to this result.

The first idea appeals to some background important facts about admiring. One such fact is that it is the kind of state to motivate you to emulate the people that you admire. That fact is a reason to be such that if you admire anyone, you only admire people who it would not be a bad idea to emulate. Moreover, this is a reason that you have, whether you admire anyone or not. Now, if you just don’t ever happen to go in for admiring – if you’re just not prone to that attitude and you don’t ever think about it – then this reason doesn’t affect you very much. But if you *are* engaged in admiring, then given this reason, you have derivative reasons to only admire Max if it would not be a bad idea to emulate him, to only admire Kat if it would not be a bad idea to

---

<sup>7</sup> As noted by Rosen [unpublished].

emulate her, and so on. On this picture, these reasons are derivative reasons which are triggered by the fact that you are engaged in admiring in the first place. So they are shared by anyone who is engaged in admiring, and hence are the right kind of reasons for admiration. And intuitively, they *are* the right kind of reasons for admiration. So that sounds right.

We can get a similar result in a different way, by taking a detour through the alethic view from part 2. If admiration is an attitude which represents its objects as being in a certain way, and if there is a standing reason not to have false mental representations of a certain kind – including the kind involved in belief, but also whatever kind is involved in admiration – then we could take the view that having the attitude of admiration triggers these reasons to not have false representations, by giving you reasons to not admire people who lack the feature that admiration represents people as having (whatever that is: see proponents of the alethic view for elaboration).

I'm not sure, exactly, which of these two ways of implementing my solution for the right kind of reasons to admire is correct – maybe both are. The main lesson that I would like to draw, however, is that there are multiple ways in which it could turn out that an activity gives rise to shared reasons, and hence to a shared standard of correctness. My solution to the Wrong Kind of Reasons problem does not turn on any particular story about *how* an activity comes to give rise to a shared set of reasons – only on whether it turns out that activities which are governed by standards of correctness also give rise to shared sets of reasons.

#### **4 confirmation**

I now take myself to have sketched in at least a strong outline what the highly general solution to the Wrong Kind of Reasons problem might look like. Wrong kinds of reasons with respect to any given activity are ones that are not in the shared set that the activity gives rise to, and hence which do not contribute to the standard of correctness for that activity. There is more work to be done in elaborating and testing this view, but for now I'll reserve that for another occasion. In this section I'm going to seek collateral confirmation for this diagnosis by showing that it straightforwardly leads to answers to problems for FA accounts of 'good for' and attributive 'good' which do not on their face of it appear to be special cases of the Wrong Kind of Reasons problem.



*'good for' and attributive 'good'*

In section I I gave a sketch of what an FA account would look like for at least one of the important senses of 'good' and of 'better'. But the sense for which I gave an FA account is not the only sense of 'good'. For a simple demonstration why, compare the following two sentences (from our original list):<sup>8</sup>

- 1 It is better for him to talk than for him to stay silent.
- 2 It is better for him to talk than to stay silent.

These two sentences look very similar. But in fact they do not say the same thing, because the word 'for' is doing different things in each sentence. Sentence 1 says that the situation in which he talks is better than the situation in which he stays silent. But sentence 2 does not say that. It says that the situation in which he talks is better *for him* than the situation in which he stays silent.

This difference in the two sentences is manifested in their intonational structure and also in their syntax. In sentence 1, it is appropriate for there to be a break in the intonational contour between 'better' and 'for'. Whereas in sentence 2, the corresponding break should come between 'him' and 'to talk'. This corresponds to the fact that in sentence 1, both occurrences of 'for' are *complementizers* like 'that' and 'whether', whereas the 'for' in 2 is merely a homonym; it designates a distinct argument place for 'good', as in 'it's easy for him to lift 300 lbs'. The sense of 'better' that appears in sentence 2, and which requires a subject for an argument, in addition to two propositional arguments, is distinct from the 'better' in sentence 1. For simplicity, I'll just call it 'better for', and call the corresponding sense of 'good' 'good for'.

In addition to the senses so far distinguished, it is also important to distinguish the class of *attributive* uses of 'good'.<sup>9</sup> These include, paradigmatically, sentences like 'she is a good skier'. The most important fact about such sentences is that they are not equivalent to 'she is good and she is a skier'. A fair bit of ink has been spilled over how these sentences are related to sentences like 'she is good' and 'she is good at skiing', but for our purposes we can stick with the paradigm cases.<sup>10</sup> In the paradigm cases of sentences like 'she is a good skier', 'good' appears to require two arguments: the thing which is being said to be a good K, and the value of 'K' for which it is being said to be a good K. This argument structure is different from the argument structure both for the original

---

<sup>8</sup> I am indebted in my thinking about what follows to unpublished work by and discussions with USC graduate student Robert Shanklin.

<sup>9</sup> See, for example, Geach [1956].

<sup>10</sup> For some of that spilled ink, see Ziff [1960], Larson [1998], Szabo [2001], Reimer [2001] and Soames [2002].

sense of ‘good’ discussed in section I, and from the argument structure for ‘good for’. So on the face of it, it is a different sense of ‘good’ and will require a different account (though compare Ziff [1960] and Finlay [2004], [unpublished], who both postulate semantic uniformity).

For our purposes here we need take no stand on whether the three senses of ‘good’ so far distinguished exhaust its senses, and allow us to account for the entire range of uses of the word ‘good’ other than in idiomatic constructions like ‘good bye’. That may be, and it may not be. The important point for us is that they are, in fact, different.

Now, nothing about the fact that we offer an FA account of *one* of these three senses of ‘good’ requires us to defend an FA account of either of the others. It could be, for example, that the account outlined in section I is correct, but that ‘good for’ is to be understood in terms of an independent concept of health, and attributive ‘good’ is to be analyzed in terms of a natural concept of function (compare Foot [2001]). But that would leave the striking question of why we use the same word, ‘good’, for all of these things, and in a way that exhibits at least a certain amount of cross-linguistic robustness. It is my own view that one of the major virtues of the FA approach is that it can be applied to give plausible accounts of *each* of these senses of ‘good’, and hence can provide us with a picture of what they have in common, despite their different argument structure, and without forcing us to try to analyze them in terms of each other.

At any rate, Chris Heathwood [2008] has challenged FA theorists to defend a FA account of ‘good for’ or explain why it should be right for one sense of ‘good’ but not another, and a very similar challenge could be posed for attributive ‘good’. In the next two sections I’ll look at a *prima facie* problem for FA accounts of each, and show how my solution to the Wrong Kind of Reasons problem makes them go away.

*the prospering of the wicked and heathwood’s ‘friendly fix’*

A natural objection to FA accounts of ‘good for’ is what Chris Heathwood [2008] calls the problem of the *prospering of the wicked*. It is good for Ted Bundy to live a long and prosperous healthy life. By this of course, I mean not that it is good that he does, but that it is good *for him* that he does. An FA account of ‘good for’ would therefore seem to require that there are reasons to have some positive attitude toward Ted Bundy’s living a long and prosperous life. But there do not seem to be reasons to have a positive attitude toward Ted Bundy’s living a long and prosperous life. Ted Bundy, after all, brutally murdered something over 30 women by bludgeoning and strangulation, many of whose heads he removed with a hacksaw, and whose bodies he saved for

long stretches of time in order to indulge his necrophilia. It goes without saying that he was a very bad man, and not deserving of a long and prosperous life. This does not seem like something that there is any sort of significant reason to want, or to have any other positive attitude towards, for that matter.

This way of putting the problem turns on the intuition that there is no reason to have a positive attitude toward Ted Bundy living a long and prosperous life. But negative intuitions about reasons are problematic.<sup>11</sup> Fortunately, we can put the problem even more acutely in a different way. Consider the comparative claim that it is better for Ted Bundy to have a long and prosperous life than to be sent to prison, but it is better for Ted Bundy to be sent to prison than for Ted Bundy to have a long and prosperous life. (That is, it is better *for him* that he lives a long and prosperous life, but better *that he is sent to prison*.) Now, assuming the FA account of ‘good’, it follows that the reasons (of the right kind) to prefer that Ted Bundy is sent to prison outweigh the reasons (of the right kind) to prefer that he lives a long and prosperous life. But according to an FA account of ‘good for’, the reasons (of the right kind) to prefer that Ted Bundy lives a long and prosperous life must outweigh the reasons (of the right kind) to prefer that Ted Bundy is sent to prison. And so without some modification, we can’t get things both ways. The FA account of ‘good’ seems to require precisely the reverse ordering on the right kind of reasons as is required by the FA account of ‘good for’. So something needs to budge.

Heathwood treats this, effectively, as a different problem than the Wrong Kind of Reasons problem, and on the face of it, it is. His suggestion about how proponents of an FA account of ‘good for’ should solve it, however, is that they should restrict the relevant reasons to the reasons of the *subject*. What is good for Ted Bundy, Heathwood suggests, should turn only on what there are reasons for *Ted Bundy* to prefer, not on what there are reasons for the rest of us to prefer. That is why, Heathwood suggests, the FA account of ‘good for’ is not falsified by the fact that the rest of us do not, in fact, have any reasons to have a positive attitude toward Ted Bundy’s living a long and prosperous life. As long as *Ted Bundy* has these reasons, everything is alright.

Now, I think that Heathwood’s constructive suggestion about how to solve the problem of the prospering of the wicked is a Trojan Horse. And where I come from, we have some experience with those. The most obvious problem with Heathwood’s suggestion, is that it limits the FA account of ‘good for’ to only be able to deal with what is good for the sort of things which can

---

<sup>11</sup> See especially chapter 5 of Schroeder [2007b].

have reasons. But unfortunately, things can be good or bad for trees, but trees don't have reasons. I'm also inclined to think that things can be good or bad for the ozone layer or for your complexion, even though the ozone layer doesn't have reasons and neither does your complexion, but the example of trees is good enough for my purposes. An FA account of 'good for' should aspire to be general enough to deal with what is good for trees, but Heathwood's 'friendly fix' would steer us out of being able to do that. I also think that there is a second problem with accepting Heathwood's fix, which the non-trivial work it does in his subsequent argument, but I'll come back to that. Suffice it to say that the fact that some things are good for trees is grounds to be suspicious of Heathwood's friendly suggestion.

Now, like Heathwood's suggestion, the solution to the Wrong Kind of Reasons problem that I've been peddling in this paper also tells us why there don't need to be reasons for *us* to have a positive attitude toward Ted Bundy's living a long and prosperous life, in order for it to be good for him to live a long and prosperous life. That is because the right kind of reasons don't need to be shared by someone who is not engaged in the relevant activity. So if the kind of 'activity' that is implicated by the positive attitude that figures in the FA account of 'good for' is one that most of us are not participating in, and which there is no particular need for us to participate in, given that Ted Bundy is beyond the moral pale, it follows that there need be no reasons for us to have that attitude. In fact, as I've understood the FA approach here, there needn't even be any reasons for *anyone* to have the attitude, if there doesn't happen to be anyone who is engaged in the relevant activity. There merely has to be a fact about what reasons *would* be shared by necessarily anyone engaged in that activity, and just because they are so engaged.

So what activity could make this true? Well, consider the activity of watching out for Ted Bundy. If Ted Bundy is the only one watching out for Ted Bundy, then only Ted Bundy's reasons will figure into the question of what is good for him – which gets us out of the problem of the prospering of the wicked. And in actual matter of fact, most of us are not engaged in the activity of watching out for Ted Bundy, and rightly so – he is beyond the moral pale. Of course, his own mother might be watching out for Ted Bundy – and in that case we would expect her to also have the reasons which bear on what is good for him. More would need to be said about the activity of watching out for Ted Bundy in order to make good on the claim that it does not need to be itself understood in terms of wanting what is good for Ted Bundy. I don't claim to be resolving that question, here. All I claim to be doing, is showing how my approach to the Wrong Kind of Reasons problem has the potential to answer the problem of the prospering of the wicked, as well.

Moreover, I believe that this answer not only avoids the problematic conclusion that nothing is good for trees (or for the ozone layer or for your complexion), but avoids the other problematic results to which Heathwood tries to commit proponents of FA accounts of ‘good for’ in the central argument of his paper. In that central argument, Heathwood contrasts the fact that what is good for a person stays the same over time, with the fact that time bias in preferences is both normal and rational. It is normal and rational to prefer slightly more pain at a somewhat later date, but it is not better for one to have slightly more pain at a somewhat later date. Hence, Heathwood rules (and I am drastically simplifying his argument, here), the FA account of ‘good for’, because it is tied to what the agent herself has reason to prefer, is forced to get the wrong results.

I hold, in contrast, that whatever reasons contribute to the rationality of time bias in preferences are of the wrong kind to contribute to the matter of what is good for one. And my theory predicts this. According to my solution to the Wrong Kind of Reasons problem, the right kind of reason needs to be shared by *anyone* who is engaged in the relevant activity. But the rationality of time bias is first-personal. It may be rational for *me* to have a bias toward the near with respect to my future pains, but it does not make sense for my mother or spouse to have the equivalent bias with respect to my future pains. So whatever reasons make my own first-personal time bias rational must not be ones that are shared by anyone who is watching out for me. Hence, they are of the wrong kind to affect what is good for me.

#### *good assassins*

FA accounts of attributive ‘good’ face an analogous problem to the problem of the prospering of the wicked. It is this: some things are bad. For example, thieves and assassins are bad. If Max is a better assassin than Jax, then he is more effective at killing people and getting away with it. But that is not a good trait to have! It’s a bad thing, for there to be good assassins floating around. Yet if an FA account of attributive ‘good’ is to work, then there must be reasons (of the right kind) to have a more positive attitude (of some kind) toward Max than toward Jax, if he is really the better assassin. Are there really such reasons?

The answer is, I think: yes, if we’re careful enough about what the relevant activity is. If you’re in the market for an assassin, then you must be trying to get someone killed. And if you’re trying to get someone killed, then it helps to hire someone who is effective at killing people and

getting away with it. So if Max is more effective at killing people and getting away with it, then you'll have reason to choose Max, rather than Jax, for the job.

Now of course, you have all sorts of reasons not to hire an assassin in the first place – it's wrong, for starters. But nothing about my solution to the Wrong Kind of Reasons problem requires that the relevant activity be morally permissible, and intuitively, that is right – there are standards of correctness even for activities that it is impermissible to participate in. For example, there are correct and incorrect ways to go fox hunting, and correct and incorrect ways to collect human scalps, and correct and incorrect ways of accommodating the Nazis, but I don't advocate doing any of those things – in fact, just to be clear, I advocate avoiding them.

In addition to the reasons that you have not to hire an assassin in the first place, there are also moral reasons for you to at least hire the less effective killer, in order to at least mitigate the potential effects. But my account predicts that those are of the wrong kind – they are reasons you have even if you're not in the market for an assassin. So all told, among all the shared reasons that arise from the activity of shopping for assassins, the reasons pretty clearly come down in favor of Max over Jax. And that's right – for Max is, after all, a better assassin than Jax.

This account, like the one sketched in the last subsection, subsumes this new problem under the solution to the Wrong Kind of Reasons problem. I'm inclined to think that this is a bonus – additional confirmation that that solution is at least on the right track.

## 5 application

If my solution to the Wrong Kind of Reasons problem *is* on the right track, however, then I think that it poses a problem for a certain kind of strategy for making sense of so-called 'agent-relative value'. In this final section I'll briefly explain this issue and how it is affected by the foregoing.

*is 'agent-relative' value a kind of value?*

Agent-Relative Teleology is the moral theory according to which every agent ought always to do that action which would bring about the most good-relative-to her, where 'good-relative-to' picks out a relation called *agent-relative value*, distinct from all of those so far described, here. The original idea behind Agent-Relative Teleology was to generalize on consequentialism in a way that makes it possible to accommodate agent-centered constraints, special obligations, and agent-centered options – all traditionally taken to be important features of ordinary moral thinking that are ruled

out by consequentialist theories – while still retaining the other attractive features of consequentialism. I won't rehearse that whole dialectic, here.<sup>12</sup>

I've argued elsewhere (Schroeder [2007a]) both (1) that agent-relative value is a technical notion, of which we have no pre-theoretical grasp, and (2) that the attractions of Agent-Relative Teleology turn on it's turning out that agent-relative value nevertheless has a lot in common with genuine value. In other words, Agent-Relative Teleology is more attractive than its ordinary deontological competitors only if the relation picked out by 'good-relative-to' deserves its name, because it has something important in common with the property or relation picked out by 'good'. But given that 'good-relative-to' is a technical term that was introduced in order to make a theory work, it is hard to see how it *could* pick out something that is so similar to what is picked out by 'good'. Hence, the challenge goes, we shouldn't be so sure that Agent-Relative Teleology is such an attractive view after all.

*the ewing-garcia-smith-portmore proposal and why it's wrong*

Now, by far the most promising attempts to argue that agent-relative value *does* have a lot in common with ordinary value have turned on the idea that both 'good' and 'good-relative-to' need to receive an FA analysis. According to the main proponents of this strategy,<sup>13</sup> something is agent-neutrally good when there is reason for everyone to desire it, and agent-relatively good relative to Max (for example) when there is reason for Max to desire it. According to Garcia [1986] and Smith [2003], for example, one of the things that we *find out* when we find out that the FA account of 'good' is on the right track, is that 'good' is agent-relative (though see Schroeder [2007a] for serious problems with this view).

Whether or not Smith's stronger claim is correct, however, if agent-relative value differs from agent-neutral value in this way, then it is easy to see what they have in common. Given how similar agent-relative value is to agent-neutral value, on this picture, Agent-Relative Teleology, which appeals to agent-relative value, and ordinary consequentialism, which appeals to agent-neutral value, have an awful lot in common. Hence, it makes sense, on this picture of what agent-

---

<sup>12</sup> See Williams [1973], Scheffler [1983], Sen [1982] and [1983], Dreier [1993], Bykvist [1996], Smith [2003], Louise [2004], Portmore [2005], and Schroeder [2007a] for further discussion of these issues.

<sup>13</sup> See Garcia [1986] and Smith [unpublished], and compare Sidgwick [1907] and Ewing [1947]. Sidgwick and Ewing were thinking about 'good for', not about the technical notion of 'good-relative-to'. Doug Portmore [2007] has advocated a similar line of reasoning to that discussed in this section, but Portmore's view is more subtle than Garcia's or Smith's and he makes no claims about agent-relative value as such.

relative value is, to expect that Agent-Relative Teleology would retain many of the advantages of ordinary consequentialism, as Agent-Relative Teleologists are prone to claim that it does.

Unfortunately, however, not only does this idea suffer from problems of its own,<sup>14</sup> my solution to the Wrong Kind of Reasons problem rules out a picture like this one. On my picture, ‘good’, ‘good for’ and attributive ‘good’ have in common that they all receive an FA analysis, and hence are to be understood in terms of the standards of correctness on some kind of activity. Standards of correctness, moreover, are fixed by the shared reasons that an activity gives rise to. On this picture, Agent-Relative Teleologists can go ahead and define ‘ $x$  is good-relative-to  $A$ ’ to mean ‘ $x$  is what  $A$  has reason to desire’, but that won’t make ‘good-relative-to’ analogous to any of ‘good’, ‘good for’, or attributive ‘good’, and it won’t allow for what I claim is the most promising solution to the Wrong Kind of Reasons problem.

#### *closing and further questions*

In this paper I’ve tried to advance a distinctive approach to the Wrong Kind of Reasons problem characterized by its broad generality, and to apply it to some related questions. I’ve tried to argue that the Wrong Kind of Reasons problem is everyone’s problem – or near enough – and that we should expect an adequate answer to work by distinguishing between reasons of the ‘right’ kind, and reasons of the ‘wrong’ kind. I’ve argued that very similar-looking problems arise in any domain governed by a standard of correctness, and hence are not the province specifically of mental states or attitudes. My own account aspires, in contrast, to be perfectly general in just these ways.

The account that I’ve offered has not been spelled out with exact precision; in each particular case it leaves important questions about how, exactly, we are to understand each of the relevant activities, who should count as engaged in them, and how those activities give rise to a distinctive set of shared reasons. I don’t pretend to have discharged those obligations; instead I

---

<sup>14</sup> Schroeder [2007a] offers an argument I think is decisive against the Garcia-Smith version of this view, according to which the good-relative-to  $x$  is what there are reasons for  $x$  to desire and the (agent-neutral) good is what there are reasons for everyone to desire. Jussi Suikkanen [unpublished] offers a revised version of the Garcia-Smith view which takes the same view of agent-relative value but instead of the universal-quantifier-involving version of the FA account of agent-neutral value, he advocates an ‘ideal observer’ FA account, according to which the (agent-neutral) good is what there are reasons for an ideal impartial observer to desire. Suikkanen’s view escapes my earlier argument but preserves the spirit of the Garcia-Smith move as a response to the challenge in Schroeder [2007a]. (It does not, however, allow Suikkanen to avail himself of the account of the conflict between neutral and relative value in Smith [2003].) The problem presented here applies even to Suikkanen’s account, however, so it is more general than my earlier argument.



settled for circumstantial evidence that any description of an activity that is rich enough to make us confident that that activity, so conceived, is really governed by a standard of correctness, will be rich enough to make us confident that it gives rise to a shared set of reasons, and for illustrating a few different ways in which this can happen.

The account I've offered has also been silent on at least one important issue that has played an important role in discussions of the Wrong Kind of Reasons problem in the literature. This is that in the case of reasons for attitudes like belief, intention and admiration, there is a significant asymmetry between the ease with which we can have these attitudes for the 'right' kind of reasons and for the 'wrong' kind of reasons. In fact, in describing it this way I significantly understate the problem; by many philosophers' accounts, it is *impossible* to believe, intend, or admire for the wrong kind of reason. Nothing I have said explains why this should be so – in fact, since there *is no* such asymmetry between the right and wrong kind of reasons in the cases of tying knots, executing the Lucena endgame, and setting the table for a White House state dinner, it is a *good* thing that my account does not explain this asymmetry, which arises only in a more limited range of cases.

My hope is that if my account of the right kind of reasons is on the right track, then the explanation for the asymmetry in ease for which we can have attitudes for the right and wrong kind of reasons will fall out from my account of the general distinction between the right and wrong kind of reasons, together with a plausible restrictive condition which applies only in the cases of reasons for attitudes like belief, intention, and admiring. Here I can only offer a conjecture: someone who lays one piece of rope across another is only contingently, if at all, engaged in the activity of tying knots, someone who moves their rook to the third rank under certain circumstances is only contingently, if at all, engaged in the activity of executing the Lucena endgame position, and someone who places a salad fork next to a soup spoon is only contingently, if at all, engaged in the activity of setting the table for a White House state dinner. But someone who believes that *p* is necessarily engaged in the activity of believing, someone who intends to do *A* is necessarily engaged in the activity of intending, and someone who admires Max is necessarily engaged in the activity of admiring. This important difference between the cases may play an important role in the ultimate explanation of the important asymmetric phenomena which arise only in the attitude-oriented cases. But that is only a conjecture, and a gesture toward problems for future research.

Finally, though versions of them have been on display here and I've tried to hint at the outline of how I think they should go, I do not even claim to have offered any real argument, here, in favor of FA accounts of value – of 'good', 'good for', or attributive 'good' – other than that this would explain what they have in common with one another. So all in all, I've left many things unresolved. But what I *do* hope to have accomplished, is to have said enough to have illustrated that the approach I'm offering is general, that it has the potential to help with other problems for Fitting Attitudes views, and that whether I am right has further implications for related issues in moral theory.<sup>15</sup>

#### references

- Bykvist, Krister [1996]. 'Utilitarian Deontologies? On Preference Utilitarianism and Agent-Relative Value.' *Theoria* 62(1): 124-143.
- D'Arms, Justin, and Daniel Jacobson [2000a]. 'The Moralistic Fallacy.' *Philosophy and Phenomenological Research* 41(1): 65-90.
- \_\_\_\_\_ [2000b]. 'Sentiment and Value.' *Ethics* 110(4): 722-748.
- Dreier, James [1993]. 'The Structure of Normative Theories.' *Monist* 76(1): 22-40.
- Enoch, David [2004]. 'Why Idealize?' *Ethics* 114(4): 759-787.
- Ewing, A.C. [1947]. *The Definition of Good*. London: Routledge.
- Finlay, Stephen [2004]. 'The Conversational Practicality of Value Judgment.' *Journal of Ethics* 8: 205-223.
- \_\_\_\_\_ [unpublished]. Unpublished book project.
- Foot, Philippa [2001]. *Natural Goodness*. Oxford: Oxford University Press.
- Garcia, J.L.A. [1986]. 'Evaluator Relativity and the Theory of Value.' *Mind* 95(2): 242-245.
- Geach, Peter [1956]. 'Good and Evil.' *Analysis* 17(1): 33-42.
- Heathwood, Chris [2008]. 'Fitting Attitudes and Welfare.' Forthcoming in *Oxford Studies in Metaethics*.

---

<sup>15</sup> Special thanks to Gideon Rosen for conversations about this topic stretching over seven years, as well as to Chris Heathwood, Doug Portmore, Russ Shafer-Landau, Stephen Finlay, Robert Shanklin, Ralph Wedgwood, David Enoch, Hagit Benbaji, Yitzhak Benbaji, to an audience in the Moral Psychology and Moral Realism research group at the Institute for Advanced Studies at the Hebrew University in Jerusalem. Time for work on this project was made possible by the joint generosity of the Institute for Advanced Studies at the Hebrew University and of the USC College of Letters, Arts, and Sciences.

- Hieronymi, Pamela [2005]. 'The Wrong Kind of Reasons.' *Journal of Philosophy*. 102(9): 437-457.
- Kavka, Gregory [1983]. 'The Toxin Puzzle.' *Analysis* 43(1): 33-36.
- Larson, Richard [1998]. 'Events and Modification in Nominals.' In D. Strolovitch and A. Lawson, eds., *Proceedings from Semantics and Linguistic Theory (SALT) XI*. Ithaca: CLC Publications, Dept. of Linguistics, Cornell University.
- Louise, Jennie [2004]. 'Relativity of Value and the Consequentialist Umbrella.' *Philosophical Quarterly* 54: 518-536.
- Mill, John Stuart [1863]. *Utilitarianism*. London: Parker, Son, and Bourn.
- Olson, Jonas [2004]. 'Buck-Passing and the Wrong Kind of Reasons.' *The Philosophical Quarterly* 54(3): 295-300.
- Parfit, Derek [2001]. In Dan Egonson, Jonas Josefson, Björn Petterson, and Toni Ronnow-Rasmussen, eds., *Exploring Practical Philosophy: From Action to Values*. Aldershot: Ashgate, 17-39.
- Piller, Christian [2001]. 'Normative Practical Reasoning.' *Proceedings of the Aristotelian Society* 75, supp. v. I, 195-216.
- \_\_\_\_\_ [2006]. 'Content-Related and Attitude-Related Reasons for Preferences.' *Philosophy* 81(1): 155-181.
- Portmore, Doug [2005]. 'Combining Teleological Ethics with Evaluator Relativism: A Promising Result.' *Pacific Philosophical Quarterly* 86(1): 95-113.
- \_\_\_\_\_ [2007]. 'Consequentializing Moral Theories.' *Pacific Philosophical Quarterly* 88(1): 39-73.
- Rabinowicz, Wlodek, and Toni Ronnow-Rasmussen [2004]. 'The Strike of the Demon: On Fitting Pro-Attitudes and Value.' *Ethics* 114(3): 391-423.
- Reimer, Marga [2002]. 'Do Adjectives Conform to Compositionality?' *Noûs* 36: 183-196.
- Rosen, Gideon [unpublished]. 'Some Basic Questions in the Theory of Responsibility.' Unpublished paper.
- Ross, W.D. [1930] (2002). *The Right and the Good*. Reprinted with an introduction by Phillip Stratton-Lake. Oxford: Oxford University Press.
- Scanlon, T.M. [1998]. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Scheffler, Samuel [1983]. *The Rejection of Consequentialism*. Oxford: Oxford University Press.
- Schroeder, Mark [2007a]. 'Teleology, Agent-Relative Value, and 'Good'.' *Ethics* 117(2): 265-295.
- \_\_\_\_\_ [2007b]. *Slaves of the Passions*. Oxford: Oxford University Press.
- Sen, Amartya [1982]. 'Rights and Agency.' *Philosophy and Public Affairs* 11(1): 3-39.

- \_\_\_\_\_ [1983]. 'Evaluator Relativity and Consequential Evaluation.' *Philosophy and Public Affairs* 12(1): 113-132.
- Shanklin, Robert [unpublished]. 'Experiencer Adjectives and 'Good'.' Unpublished dissertation proposal, University of Southern California, 2008.
- Sidgwick, Henry [1907]. *The Methods of Ethics*, 7<sup>th</sup> edition. London: MacMillan.
- Smith, Michael [2003]. 'Neutral and Relative Value after Moore.' *Ethics* 113: 576-598.
- \_\_\_\_\_ [unpublished]. 'Two Kinds of Consequentialism.' Unpublished draft manuscript.
- Soames, Scott [2002]. *Philosophical Analysis in the Twentieth Century Volume 2: The Age of Meaning*. Princeton: Princeton University Press.
- Suikkanen, Jussi [unpublished]. 'Consequentialism and the Good-Relative-To: A Reply to Mark Schroeder.' Unpublished draft manuscript.
- Szabo, Zoltan Gendler [2001]. 'Adjectives in Context.' In I. Kenesei and R.M. Harnish, eds., *Perspectives on Semantics, Pragmatics, and Discourse*. Amsterdam: John Benjamins, 119-146.
- Williams, Bernard [1973]. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Ziff, Paul [1960]. *Semantic Analysis*. Ithaca: Cornell University Press.