
REASONS WRONG AND RIGHT

BY

NATHANIEL SHARADIN

Abstract: The fact that someone is generous is a reason to admire them. The fact that someone will pay you to admire them is also a reason to admire them. But there is a difference in kind between these two reasons: the former seems to be the ‘right’ kind of reason to admire, whereas the latter seems to be the ‘wrong’ kind of reason to admire. The Wrong Kind of Reasons Problem is the problem of explaining the difference between the ‘right’ and the ‘wrong’ kind of reasons wherever it appears. In this article I argue that two recent proposals for solving the Wrong Kind of Reasons Problem do not work. I then offer an alternative solution that provides a unified, systematic explanation of the difference between the two kinds of reasons.

1. Introduction: two kinds of reasons

One thing philosophers agree on is that, whatever else is true of reasons, reasons *count in favor* of what they are reasons for.¹ When it comes to *attitudes* like belief and admiration, something is a reason when it counts in favor of the attitude. Among reasons for attitudes, we can distinguish between intuitively *good* reasons and intuitively *bad* reasons, where intuitive quality is a matter of strength. For instance, testimony from a reliable source to the effect that the concert begins at eight is in this sense a good reason, i.e. a relatively *strong* reason, to believe the concert begins at eight. By contrast, my dim recollection of last year’s concert beginning at eight is a bad reason, i.e. a relatively *weak* reason to believe the concert begins at eight. It is an interesting question what accounts for the difference in strength among reasons for attitudes.² But in this article I am interested in a different difference among reasons, one that is orthogonal to difference in strength. To illustrate the difference I am interested in, consider the following two reasons for believing the concert begins at eight:

Memory: I dimly recall last year’s concert beginning at eight.

Criminals: Criminals will torture my family unless I believe the concert begins at eight.

As we just noted, Memory is a relatively *weak* reason for belief. By contrast, Criminals seems like a relatively *strong* reason for belief. After all, the fact that criminals will torture my family unless I believe the concert begins at eight counts *quite strongly indeed* in favor of so believing. But despite being a relatively strong reason, there is something funny, odd, or, as we might put it, not quite kosher, about the reason constituted by Criminals. Memory and Criminals seem to differ not just in strength, but also in the *kind* of reasons they are: What accounts for *this* difference between reasons for belief?

One natural reaction to have to cases such as this is to distinguish between *epistemic* and *non-epistemic* reasons for belief along the following lines: epistemic reasons for belief bear on the truth of the thing believed, whereas non-epistemic reasons do not. Memory, despite being a relatively weak reason, is an epistemic reason for belief, whereas Criminals, however strong a reason it is, is a non-epistemic reason for belief. But natural as it is, this way of accounting for the difference between Memory and Criminals cannot be the whole story. This is for two reasons. First, we shall still want to know why it is that, compared to epistemic reasons for belief, *non-epistemic* reasons have the distinctively odd flavor they do. At the very least, then, we are owed a story about that. Second, the *very same oddness* among reasons seems to show up in reasons for attitudes other than belief. Consider the following two reasons for admiring a person, Lara:

Generous: Lara is generous.

Benefactor: Lara's benefactor will pay me to admire her.

Regardless of whatever difference there might be in the relative strengths of Generous and Benefactor, there is this difference between them: compared to Generous, there is something funny, odd, not quite kosher, about Benefactor as a reason to admire Lara. Moreover, the oddness exhibited by Benefactor seems to be *exactly the same* oddness exhibited by non-epistemic reasons such as Criminals in the case of belief. In addition to a story about why non-epistemic reasons for belief have the odd flavor they do, then, we are also owed a story about what unifies the oddness exhibited by those reasons and the oddness exhibited by some of the reasons for attitudes like admiration. Simply describing the difference in terms of epistemic and non-epistemic reasons for belief will not do the trick, because the very same difference seems to show up in cases where the distinction between epistemic and non-epistemic reasons is inapt, e.g. in the case of attitudes such as admiration.

What we have here is a phenomenon that appears across a range of attitudes for which there can be reasons, including belief and admiration, but also including blame, envy, fear, love, and desire. The phenomenon is

usually labeled the phenomenon of the 'Wrong Kind of Reasons.'³ The idea is that reasons like Memory and Generous are the 'right,' whereas what I have been calling 'odd' or 'not quite kosher' reasons like Criminals and Benefactor are the 'wrong,' kind of reasons for their respective attitudes. This label for the phenomenon is potentially misleading. It is potentially misleading because the label 'wrong' suggests there is something *defective* or somehow *worse* about this sort of reason. But as we have already seen, Criminals and Benefactor seem to be quite good reasons, in the sense of being relatively strong reasons, for their respective attitudes. This is not to say there is nothing odd about such reasons. This is to say that whether their oddness amounts to a defect in them will depend on what account we give of the phenomenon. With this caveat in mind, I will follow the convention of calling reasons like Memory and Kind the 'right' kind of reasons and reasons like Criminals and Benefactor the 'wrong' kind of reasons. So understood, the phenomenon of the Wrong Kind of Reasons presents a philosophical *problem*: The Wrong Kind of Reasons Problem is the problem of giving a systematic account of the difference between the right and the wrong kind of reasons for attitudes.

You might worry that the problem has no solution, because there is not anything that unifies the right and the wrong kind of reasons for diverse attitudes such as belief and admiration. You might worry that for each of these different attitudes, there is an intuitive distinction between two different kinds of reason, but that all these different dichotomies are all different from one another. Here is an analogy: perhaps being a right kind of reason for belief as opposed to a wrong kind of reason for belief is no more the same kind of thing as being a right kind of reason for admiration as opposed to a wrong kind of reason for admiration than being a golden parachute (for a CEO) as opposed to a Labrador parachute (a certain kind of clause in a CEO's contract that first became common on the island of Labrador) is the same kind of thing as being a golden retriever as opposed to a Labrador retriever.⁴ The proper response to this worry comes in two steps.

The first step is to point out that, whatever attitude we are concerned with, the right kind of reasons share some distinguishing features.⁵ One such feature is that the right kind of reasons exhibit a motivational asymmetry to the wrong kind of reasons: it is in general easier to get oneself to believe or admire for the right kind of reason than for a wrong kind of reason.⁶ Another feature is that the right kind of reasons, but not the wrong kind of reasons, seem to bear on the rationality of the attitude for which they are reasons in a distinctive way: the right kind of reasons, but not the wrong kind of reasons, seem to bear, in a way that is difficult to state precisely, on whether an attitude is rational as an instance of the kind of attitude it is. This is a rough way of putting the idea, but that is part of the point of the response to the worry, since the second step in the response is a conditional appeal to explanatory unity: if we can explain, in a unified, systematic way, why it is

that the right and the wrong kind of reasons seem to have the features they do, and what these features actually amount to, then we should. Doing so is part of the task of giving a unified, systematic account of the nature of reasons in general. The project of this article, in part, is to show that we *can* give such an explanation: that we can contribute to our understanding of the nature of reasons in general by giving a unified systematic account of the difference between the right and the wrong kind of reasons for attitudes. I am therefore offering a promissory note, cashable on the condition that we can explain the intuitive distinction in a systematic, unified way.

Before presenting my explanation of that distinction, I am going to argue that two recent attempts to account for the difference between the right and the wrong kind of reasons fail. Despite failing, each attempt contains important insights into the problem. So it is worth spending time with these accounts in order to see whether their insights can be incorporated into an alternative solution. As I will argue, the clue to explaining the difference between the right and the wrong kinds of reasons for attitudes lies in noticing that, despite what we might have thought, the distinction between the right and the wrong kind of reasons does not arise for all attitudes for which there can be reasons. As we will see, this is *prima facie* puzzling. I will argue that what explains this puzzle also provides the resources for explaining the distinction between the two kinds of reasons where it does arise. And, as I will argue, the resulting account incorporates the insights of the two failed competitor accounts. More on this later. First, let me try and convince you that two recent attempts to account for the difference between the right and the wrong kind of reasons do not work.

2. *Two failed solutions*

HIERONYMI

Pamela Hieronymi has recently suggested the following way of accounting for the difference between the right and the wrong kind of reasons. According to Hieronymi, the clue to explaining the difference lies in noticing that certain attitudes, belief among them, are what she calls ‘commitment-constituted.’⁷ A commitment-constituted attitude is one for which there is a question the answering of which amounts to forming that attitude. It is easy to see how this works for the attitude of belief. Recall the example from above of the belief that the concert begins at eight. Plausibly, affirmatively answering for oneself the question ‘whether the concert begins at eight’ simply amounts to forming a belief that the concert begins at eight.⁸ In other words, settling for oneself the question ‘whether the concert begins at eight’ *just is* forming a belief about whether the concert begins at eight. Having noticed this feature of commitment-constituted attitudes, Hieronymi points out that there is a corresponding

difference in reasons for such attitudes: some reasons for these attitudes bear on a question the answering of which amounts to forming the attitude, and some do not. In our example, Memory bears on the question 'whether the concert begins at eight' whereas Criminals does not. Intuitively, Criminals bears on a different question, such as 'whether it would be good to believe the concert begins at eight.' Now, according to Hieronymi, this difference, the difference between bearing on a question the answering of which amounts to forming an attitude and not bearing on a question the answering of which amounts to forming an attitude, is what the difference between the right and the wrong kind of reasons for an attitude amounts to. Memory is the right kind of reason to believe the concert begins at eight because Memory bears on a question the answering of which amounts to believing the concert begins at eight, and Criminals is the wrong kind of reason because Criminals does not bear on this question. In general, then, Hieronymi's view is that:

Bears on a Question (BQ): R is the right kind of reason to φ_{mind} iff R bears on a question the answering of which amounts to φ_{mind} -ing.⁹

The problem with Hieronymi's account is this: BQ is extensionally inadequate to the phenomenon. This is because, for some attitudes, there is a difference between the right and the wrong kind of reasons for the attitude, but there is no question the answering of which amounts to forming the attitude. Thus, for these attitudes, while there are not any reasons that bear on a question the answering of which amounts to forming the attitudes (since there is no such question), there is nonetheless a difference between the right and the wrong kind of reasons for the attitudes.

Admiration is an attitude like this. There simply is not a question the answering of which amounts to forming the attitude of admiration toward some person *N* analogously to the way answering the question 'whether *P*' amounts to forming the belief that *P*. To insist otherwise is to do violence to what we think admiring someone is like: it is to collapse the distinction between thought and feeling.¹⁰ When I admire Lara, it is not as if all I do is make or assent to certain judgments about her: that she is courageous, that she is generous, and so on. It is certainly a complicated matter to say what *else* admiring Lara involves, but we can safely say this: admiring Lara involves a distinctive way of being attitudinally related to Lara that is not exhausted by being credally related to certain propositions, i.e. by having certain beliefs, or by making or assenting to certain judgments. But, quite generally, settling for oneself a question *is* exhausted by having certain beliefs, or by making or assenting to certain judgments.¹¹ But then, we can be certain there is no question such that settling for oneself that question amounts to admiring Lara, since the former is simply a matter of having certain beliefs, and the latter, whatever else it is, is not only that. But this means that there will not be any facts that bear on a question the answering

of which amounts to admiring Lara. And so there will not be any facts such that they are the right kind of reasons to admire Lara. But this is false. Recall the difference between:

Generous: Lara is generous.

Benefactor: Lara's benefactor will pay me to admire her.

Generous is clearly the *right* kind of reason to admire Lara, whereas Benefactor is clearly the *wrong* kind of reason to admire Lara. The problem is that Hieronymi's account (BQ) will not allow us to say this. That is because, according to BQ, for Generous to be the right kind of reason to admire Lara is for it to bear on a question the answering of which amounts to admiring Lara. But there is no such question. And so nothing bears on that question; and so in particular Generous does not bear on it. So, if BQ is correct, then then we are blocked from saying what it is clearly correct to say about this case.

If you are not convinced because you think, contrary to my claim, that there is a question the answering of which amounts to admiring someone, consider the general form of the argument. I claimed that (i) there can be the right and the wrong kind of reasons to admire a person and, while admiring a person may *involve* having certain beliefs, (ii) admiring a person also involves *more* than simply having certain beliefs. All that is required to show that Hieronymi's account is false is that (i) and (ii) be true *for some attitude or other*. I think (i) and (ii) are clearly both true of *admiration*. But we can just as easily pick a different attitude for which the difference between the right and wrong kind of reasons shows up but where the presence of the attitude is not guaranteed by the presence of certain beliefs; my claim is that there is at least one such attitude. The attitude of *love* is a particularly good alternate case.¹² Given the existence of attitudes for which a distinction between the right and the wrong kind of reasons for those attitudes makes sense, but for which the idea of a question the answering of which amounts to having the attitude does not, BQ is unacceptable as a solution to the Wrong Kind of Reasons Problem.

The point can be made even more forcefully than this, if we think there are sometimes reasons for evaluative attitudes that are best thought of as emotions, such as anger, or fear. This is because the difference between the right and the wrong kind of reasons will appear for these emotions just as it did for attitudes like admiration, belief, and so on. For instance, the fact that something is dangerous is a reason – intuitively, the right kind of reason – to be afraid of it. And the fact that someone will pay you a substantial amount of money to be afraid of something is a reason – intuitively, the wrong kind of reason – to be afraid of it. How could Hieronymi account for this difference? Her view would have to be that there is a question the answering of which amounts to forming the emotion of fear, and that the right kind of reasons are considerations that bear on that question. But, perhaps even more clearly

than in the case of attitudes like admiration, there is no such question: no amount of settling questions for oneself will ever amount to having an emotion such as fear. So, if we think there can be reasons to have emotions such as fear, then Hieronymi's account of the distinction between the right and the wrong kind of reasons cannot be extended to cover this distinction as it arises in cases of reasons for these emotions.¹³

SCHROEDER

Faced with similar worries with the extensional adequacy of Hieronymi's account, Mark Schroeder has recently offered a somewhat different account of the distinction between the right and the wrong kind of reasons for attitudes.¹⁴ According to Schroeder, the 'point of the distinction between the "right" and the "wrong" kind of reasons, is that only the "right" kind contribute to standards of correctness'.¹⁵ For instance, the right kind of reasons for belief seem to contribute to whether a belief is correct, and the wrong kind of reasons for belief do not. Why should this be? Schroeder's idea is that the wrong kind of reasons are *idiosyncratic* in a way that bars them from contributing to standards of correctness.¹⁶ In the case above, Criminals is a reason for *me* to believe the concert begins at eight, but not a reason for *you* – or anyone else, for that matter – to believe the concert begins at eight. Similarly, Benefactor is a reason for *me* to admire Lara, but not a reason for *you* – or anyone else, for that matter – to admire her. Intuitively, it is because these reasons are idiosyncratic that they fail to contribute to the attitudes' standards of correctness. Not only so, but also: the right kind of reasons seem to be reasons one has *simply* in virtue of having the relevant attitude, whereas the wrong kind of reasons seem to be reasons one has *also* in virtue of some other facts, such as the fact that one does not want one family to be tortured, or the fact that one wants to be paid by Lara's benefactor. The fact that the right kind of reasons are ones that one has simply in virtue of having the relevant attitude is part of what makes it the case, according to Schroeder, that the right kind of reasons are the ones that contribute to an attitude's standard of correctness. Schroeder's thought, then, is that a systematic account of the difference between the right and the wrong kind of reasons will be one that distinguishes these reasons in terms of the former, but not the latter, being reasons that are *not idiosyncratic* in these two ways: that is, they will be reasons every agent engaged in having the attitude has, and simply because she is engaged in having the attitude. Here, then, is what Schroeder suggests as a principle for distinguishing reasons of the right from reasons of the wrong kind:

Shared Reasons (SR): Relative to the attitude of φ_{mind} -ing, R is the right kind of reason to φ_{mind} iff R is a reason shared by necessarily anyone engaged in φ_{mind} -ing and just because they are so engaged.¹⁷

Intuitively, SR correctly sorts reasons like Criminals and Benefactor as the wrong kind of reasons. Recall, the idea is that the right kind of reasons contribute to an attitude's correctness because they are reasons shared by necessarily everyone engaged in having an attitude and just because they are so engaged. Criminals and Benefactor do not contribute to the correctness of believing and admiring because they are not reasons everyone engaged in believing and admiring shares. And, so, they are the wrong kind of reasons for their respective attitudes. Schroeder's proposal has some intuitive appeal. Intuitively, there is something right about the idea that the right, but not the wrong, kind of reasons contribute to an attitude's correctness. And, further, there is something intuitively right about the idea that reasons that contribute to an attitude's correctness are reasons that everyone engaged in having an attitude shares.

But because Schroeder proposes to understand correctness for attitudes *in terms* of there being shared reasons, his way of solving the problem requires that we be able to make independent sense of two claims: (i) that there is a set of reasons shared by necessarily anyone engaged in φ_{mind} -ing and just because they are so engaged (call this claim Shared Set) and (ii) that the shared set of reasons is coextensive with the set of the right kind of reasons for φ_{mind} -ing (call this claim Shared-Right Connection). The problem with Schroeder's account, I think, is that it cannot establish Shared Set and Shared-Right Connection.

Schroeder is aware that his account owes us some story about Shared Set and Shared-Right Connection, and he tries out two different strategies to establish these claims: the *background facts* strategy and the *alethic* strategy. What each of these strategies is designed to show is that there is a set of shared reasons relative to an attitude (Shared Set), and that this set of reason is equivalent to the set of the right kind of reasons for the attitude (Shared-Right Connection). If either strategy were successful, it would show that the right kind of reasons contribute to an attitude's correctness *because* they are members of a set of reasons shared by necessarily everyone engaged in having the attitude and just because they are so engaged. I will now argue that neither of Schroeder's two strategies can successfully establish both Shared Set and Shared-Right Connection. Briefly, the problem with the background facts strategy is that it cannot establish Shared Set, and the problem with the alethic strategy is that, even if it can establish Shared Set, it cannot establish Shared-Right Connection – that is, it cannot establish that the shared set of reasons is coextensive with the set of the right kind of reasons.

The background facts strategy

Schroeder's first strategy appeals to *background facts* about attitudes.¹⁸ For example, here is Schroeder discussing background facts about admiration:

One such fact is that [admiration] 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. [...] On this picture, these reasons [only to admire people who it would not be a bad idea to emulate] 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.¹⁹

The idea is that Shared Set is true for admiration because there is a shared set of reasons for emulation, and emulation naturally follows admiration: the shared set of reasons for emulation derivatively yields the shared set of reasons for admiration. Grant that there is a shared set of reasons for emulation.²⁰ This will still not do the trick to establish Shared Set for *admiration*. It is not enough to point out, as Schroeder does, that admiration is the kind of state to motivate you to emulate people you admire. For, all this shows is that admiration typically, normally, when all goes well, motivates you to emulate people you admire. But this does not show that anyone engaged in admiring a person has a reason to admire them only if it would not be a bad idea to emulate them. It would only show that if admiration always, without fail, no matter what, motivated you to emulate those you admire. But admiration does not do this. For instance, it is possible to admire someone's life-long commitment to saving the rain forests without being motivated to emulate that person. But then, since there is conceptual room to admire without also being motivated to emulate, then the reasons there are to only admire those it would not be a bad idea to emulate will not always be 'triggered' by the fact that one is engaged in admiring in the first place. In particular, these reasons will not be triggered in cases where the (admittedly typical) connection between admiring and emulating is somehow interrupted, or blocked.

You might think I am being unfair to Schroeder. Surely the connection between certain attitudes and related activities is more than merely *typical*. Emulating someone is not just typically associated with admiring them; the connection is stronger than this. I am willing to grant that the connection is stronger. But unless we are willing to strengthen the connection all the way to *necessity*, the point stands: what is required in order to show that the shared set of reasons for emulation always derivatively yields a shared set of reasons for admiration – what are supposed to be reasons of the right kind for admiration – is a necessary connection between admiration and emulation. And, I claim, however strong we think the connection between admiration and emulation is, we do not think it is a *necessary* connection.²¹

This problem with the background facts strategy generalizes. In general, there is not a *necessary* connection between someone's attitudinizing in a certain way (e.g. admiring) and that person's engaging in an intentional activity with an aim (e.g. emulation) for which, we grant, there might be a shared set

of reasons. In other words, while there might be certain activities or actions *typical* for people with certain attitudes, most attitudes – including those liable to the Wrong Kind of Reason Problem – do not come with a *necessary* connection to any actions, and so do not come with a shared set of reasons, even if certain actions do come with shared sets of reasons.²² So the background facts strategy does not work because it cannot establish Shared Set.

The alethic strategy

Schroeder's second strategy is the *alethic* strategy; here he is explaining it:

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 [...]²³

The problem with the alethic strategy is that it cannot establish Shared-Right Connection: that the shared set of reasons for an attitude is coextensive with the set of the right kind of reasons for the attitude. Consider the case of admiration. Grant that admiration represents its object as being a certain way. Grant that there is a standing reason not to have false mental representations of a certain kind, including those involved in admiration. The problem is that the reasons there are to not admire people who lack whatever features admiration represents people as having clearly do not exhaust the set of the right kind of reasons with respect to admiration. That is because they are only *negative* reasons: reasons *against* admiring people who lack certain features. But some of the right kind of reasons with respect to admiration are reasons *for* admiring certain people, not just reasons *against* admiring others. To explain: Suppose, for the sake of argument, that admiring someone involves representing them as *generous*. Now suppose Lara is *not* generous. The alethic strategy reveals why the fact that Lara is not generous is the right kind of reason *against* admiring Lara. This is because admiring Lara would be a way of falsely representing Lara as generous. But suppose Lara *is* generous. The alethic strategy does not explain why this fact is the right kind of reason to actually admire Lara. But, clearly, it is.²⁴

The case against the alethic strategy is even clearer when it comes to the attitude of belief: the reasons there are to not believe false propositions clearly do not exhaust the right kind of reasons with respect to belief. That is because they are only *negative* reasons: reasons *against* believing propositions that are false. But some of the right kind of reasons for belief are reasons *for* believing certain propositions, not just reasons *against* believing others. To explain: Suppose you receive reliable testimony that the concert

does *not* start at eight. The alethic strategy explains why this fact is the right kind of reason to *not* believe that the concert starts at eight. But now suppose you receive reliable testimony that the concert *does* start at eight. The alethic strategy does not explain why this fact is the right kind of reason to actually believe the concert starts at eight. But, clearly, it is.

The general problem with the alethic strategy is that the reasons it establishes as shared reasons are only reasons *against* having *false* mental representations. So you could try and rehabilitate the alethic strategy by extending it to include standing reasons not just against having false mental representations, but also standing reasons *for* having *true* mental representations. But that will not work either. That is because it is overwhelmingly implausible that there is such a reason. If there were a standing reason to have true mental representations, then there would be in particular a standing reason to have *beliefs* in *true* propositions, for these are paradigmatic instances of true mental representations. Then there would be a standing reason to have a belief in *any old* true proposition, no matter how trivial: there would be a reason, for instance, to have a true belief about each name and number in the phonebook. But there is no such reason. So there is not any standing reason to have true mental representations, and appealing to such a reason cannot rehabilitate the alethic strategy. So the alethic strategy does not work because it cannot establish Shared-Right Connection, that the shared set of reasons there are relative to an attitude is co-extensive with the right kind of reasons for the attitude.

At this point the fan of the alethic strategy might try to reply by discrediting our negative existential intuitions about reasons.²⁵ Elsewhere, Schroeder has tried to do precisely this.²⁶ His suggestion, briefly, is that ‘there is a reason to φ_{mind} ’ is usually elliptical for ‘there is a particularly weighty reason to φ_{mind} ’.²⁷ Without going into details: this means that our negative intuitions about the existence of particularly weak reasons cannot be trusted, because it will strike us as counterintuitive or false that there is a reason to φ_{mind} whenever the reasons for φ_{mind} -ing are sufficiently weak.²⁸ In the present context, then, the suggestion would be that the negative intuitions I appealed to above, e.g. that there is no reason to have a true belief about each name and number in the phonebook, cannot be trusted. Instead, there is a standing reason to have true mental representations, including the kind involved in admiration and belief, only it is a relatively weak reason: That is why it seemed counterintuitive that such a reason existed.

The alethic strategy that replies in this way still faces two problems. First, this reply would seem to entail that the right kind of reasons to φ_{mind} are all relatively weak reasons. To see this, recall that on the current proposal the right kind of reasons to φ_{mind} are supposed to be ‘triggered’ by the reasons there are to have true mental representations. For instance, the right kind of reasons to admire Lara, such as that she is generous, are triggered by the reasons there are to have true mental representations of the kind involved in admiration,

presumably representations to the effect that Lara is generous. We just said that the reason there is to have true mental representations about Lara is a relatively weak reason. That is what was supposed to explain our negative intuition about the existence of such a reason. But then, on the plausible assumption that facilitative connections between reasons do not contribute to the strength of a reason,²⁹ the reasons there are to admire Lara that are triggered by the reason to have true mental representations about Lara are correspondingly weak. And that seems like a mistake. The fact that Lara is generous is, in addition to being the right kind of reason to admire her, a relatively *strong* reason to admire her.

In any event, the alethic strategy faces a second, worse problem. Suppose there is a standing reason to have true mental representations, including the kind of representations involved in admiration, and suppose further that these reasons are suitably strong. The problem is that the reasons there are to have whatever true mental representations are involved in admiring Lara do not correspond to the right kind of reasons to admire Lara. That is because some of the mental representations involved in admiring Lara do not have anything to do with whether Lara is admirable, in the sense of deserving admiration, but rather have to do with whether Lara is admirable in the sense of being a suitable possible object of admiration. And only reasons for the former, and not the latter, sort of mental representation are the right kind of reasons to admire Lara. For example, suppose one of the mental representations involved in admiring Lara is the representation of Lara as a responsible agent. That is, you would not count as admiring Lara unless you had the mental representation of Lara as a responsible agent. Then, according to the account on offer, being engaged in admiration triggers reasons to have *true* mental representations about whether Lara is a responsible agent. It would follow, then, that the reasons there are to believe truly that Lara is a responsible agent are the right kind of reasons to admire Lara, since these would be reasons shared by necessarily anyone engaged in admiring Lara and just because they are so engaged. But the reasons there are for thinking that Lara is a responsible agent are not *per se* reasons (let alone reasons of the right kind) for admiring Lara.³⁰ For instance, the fact that Lara is a human adult is (at least some) reason for believing Lara is a responsible agent; but the fact that Lara is a human adult is not by itself a reason for admiring Lara, let alone a reason of the right kind for doing so. In general, the problem can be put like this: if the current suggestion is correct, the reasons there are to correctly represent the world in all the ways involved in φ_{mind} -ing are all the right kind of reasons to φ_{mind} . But that is false. For, as we have just seen, not all ways φ_{mind} -ing represents the world to be are ways that are relevant to whether φ_{mind} -ing is merited, in the sense of deserved. And thus not all reasons for correctly representing the world in the way involved in φ_{mind} -ing are the right kind of reasons to φ_{mind} . So the alethic strategy still cannot establish Shared-Right Connection.

I have just argued that Schroeder's account – SR – cannot be correct because, in order for it to work, he would have to establish both that there is a set of reasons shared by necessarily anyone engaged in *φ*_{mind}-ing and just because they are so engaged and that this set of reasons is coextensive with the set of the right kind of reasons. The problem is that neither of Schroeder's two strategies for establishing these two claims will do the trick.

3. *A puzzle and a proposal*

Where does this leave us? It leaves us without an account of the Wrong Kind of Reasons Problem: neither Hieronymi's BQ nor Schroeder's SR can successfully distinguish the *right* from the *wrong* kind of reasons for attitudes like admiration and belief. BQ was extensionally inadequate because for some attitudes, e.g. admiration, there simply is not a question the answering of which amounts to forming that attitude. And as we just saw, SR was inadequate because it relied on the idea that there are sets of reasons shared by necessarily anyone engaged in having an attitude, an idea that neither of Schroeder's two strategies could make sense of.

I do not want to make any more hay out of how Hieronymi and Schroeder's accounts go wrong, because I think each contains an important insight into the nature of the Wrong Kind of Reasons Problem. It is worth pausing over these insights for just a moment – I will return to them in some more detail below, in my concluding remarks.

Schroeder's idea, recall, is that, as he puts it, the 'point of the distinction between the "right" and the "wrong" kinds of reasons, is that only the 'right' kind contribute to standards of correctness'.³¹ The idea, then, was that *shared* sets of reasons for attitudes contribute to standards of correctness, and so it is the *shared* sets of reasons that are the right kind of reasons for an attitude. The problem with Schroeder's account, as we just saw, was that it required us to have an independent grip on the idea that there were shared sets of reasons for attitudes, an idea that neither of Schroeder's two strategies could make sense of. But despite the failure of the account, I think Schroeder's idea is basically right: the right kind of reasons are ones that contribute to standards of correctness. The problem with his view was that it required us to *derive* the standards of correctness for attitudes from the shared sets of reasons for those attitudes. Hieronymi's view also contains an important insight. Hieronymi's idea, recall, was that the right kind of reasons for an attitude are reasons that bear on a particular question. For instance, in the case of belief, her view was that the right kind of reasons for belief bear on a question the answering of which amounts to forming the belief. The problem with Hieronymi's account was that it could not be extended to cover all attitudes for which the problem arises. The case of

admiration was particularly striking: since there is no question the answering of which amounts to admiring someone, Hieronymi's account was incapable of account for the difference between the right and the wrong kind of reasons for admiration. The trouble with Hieronymi's account, I think, is not that she is wrong that the right kind of reasons bear on a particular question. The trouble is that she has got the question wrong. I am going to argue for an alternative solution to the Wrong Kind of Reasons Problem that can be thought of as combining Schroeder and Hieronymi's insights into a single account: the right kind of reasons are those that contribute to standards of correctness, and the way they do so is by bearing on a particular question, viz. the question of whether the relevant attitude is correct. This is rough. Before fleshing out the details, let me motivate my proposal by introducing a puzzle.

When I introduced the Wrong Kind of Reasons Problem for Attitudes I said that the problem of distinguishing between the right and the wrong kind of reasons was not limited to the attitudes of admiration and belief, but that the same problem arises across a range of attitudes for which there can be reasons. This way of stating the problem left open the possibility that there are attitudes for which the distinction between the right and the wrong kind of reasons does not make any sense. In fact, I think this possibility is actual. And I think focusing on cases of attitudes for which the distinction between the right and the wrong kind of reasons does *not* make any sense can help shed light on how to think about that distinction in cases where it clearly does.

Consider the attitude of imagination. Like believing, imagining is a way of being attitudinally related to a proposition, *P*. And like believing that *P*, imagining that *P* is a way of regarding *P* as true.³² Moreover, as with the attitude of belief, there can be reasons to imagine one way rather than another. For instance, consider the proposition:

Democrat (D): The Democrats will retain control of the Senate in 2014.

There can be reasons to believe D and there can be reasons to imagine D. For instance, the fact that polling data supports the truth of D is a reason to believe D, and the fact that it is part of a pleasant fantasy is a reason to imagine D. But notice: While there can be the wrong kind of reasons to believe D (that is something we are already familiar with) there cannot be the wrong kind of reasons to *imagine* D. To see this, recall how easy it was to generate reasons of the wrong kind for attitudes like belief and admiration. All we had to do was to introduce an extraordinary incentive in favor of the attitude, such as criminals threatening to torture your family, or eccentric benefactors offering payment.³³ But this method will not work for the attitude of imagination. Consider:

Pleasant: Imagining D is part of engaging in a pleasant fantasy.

Mad Billionaire: A mad billionaire will pay you a million dollars to imagine D.

Both Pleasant and Mad Billionaire are reasons to imagine D but, unlike the case of belief, the extraordinary incentives introduced in Mad Billionaire do not seem to make it a reason of the 'wrong' kind. Indeed, it is hard to see what could even count as a reason of the 'wrong' kind to imagine D: that is, while there can be reasons to imagine, imagination does not appear to be the sort of attitude for which the intuitive distinction between the 'right' and the 'wrong' kind of reasons makes sense. This is *prima facie* puzzling. It is puzzling because, in other respects, imagination seems very much like belief: both attitudes involve regarding their objects as true, both are attitudes for which there can be reasons, and so on. So: What explains why belief, but not imagination, is liable to the Wrong Kind of Reasons Problem? Put generally: What explains why some attitudes, but not others, are liable to the Wrong Kind of Reasons Problem? I think the clue to answering this question lies in noticing an important difference between attitudes like belief and imagination. Let me warm you up to that difference by way of an analogy.

Consider two visually similar objects, each of which comprises colored intersecting lines arranged in a roughly web-like structure, one of which is a paint splatter and one of which is a commuter's map of the London Underground. What, we can ask, is the difference between these two objects? There are a number. But here is one important difference: whereas we can evaluate both objects in a variety of ways, for instance in terms of their aesthetic qualities, the cost of producing each, and so on, the object that is the map of the Underground *brings with it* a particular standard of evaluation. That standard is, roughly, accuracy in conveying the (commuter's map relevant) features of the London Underground. What does it mean to say that the map of the Underground *brings with it* a particular standard of evaluation? It means that part of *what it is* to be a map of the London Underground is to be subject to this standard of evaluation. An object that we did not think of as liable to this standard would not be one that we thought of as a member of the kind 'map of the London Underground': being a series of colored intersecting lines arranged in a roughly web-like structure that is a member of the kind 'map of the London Underground' means, in part, being liable to the standard of evaluation of accurately conveying the relevant features of the London Underground.³⁴ This is not to say that we cannot similarly evaluate the paint splatter in terms of how well it accurately represents the London Underground. Perhaps by some chance the paint splatter is adequate, or even exceptional, in regards to this standard. This is to say that the paint splatter is not, simply in virtue of the kind of thing it is, the kind of thing that is liable to such evaluation, whereas the map of the Underground is so liable. And this is not to say that we cannot evaluate the map in other terms, such as whether it has artistic flair, or whether it is economical to

produce. This is to say that the map is not, simply in virtue of the kind of thing it is, the kind of thing that is liable to such evaluation.

Let us sharpen up the language. We can say that the map is subject to a *constitutive* standard of correctness involving its accuracy in representing the features of the London Underground, whereas the paint splatter, while it might be subject to some standards of correctness or other – including perhaps the standard of accuracy in representing the features of the London Underground – is not subject to any *constitutive* standard of correctness. In other words, there is no way a paint splatter should be, if it is to be a correct member of the kind ‘paint splatter’. A constitutive standard of correctness can therefore be thought of as a standard of correctness that expresses a constitutive *ideal* for the thing in question – a way the thing must be, if it is to be a correct instance *of the kind of thing it is*. In order for the map to meet the standards to which it is subject *qua map of the London Underground* it has to be accurate in representing the relevant features of the Underground. That is what it means to say that accuracy is the constitutive standard of correctness for a map of the Underground.

How does all of this relate to belief and imagination? Consider two structurally similar mental attitudes, each of which comprises an attitude of ‘regarding as true’ toward some proposition, *P*, one of which is an imagination, and one of which is a belief. What, we can ask, is the difference between these two mental attitudes? There are a number. But here is one important difference: whereas we can evaluate both attitudes in a variety of ways, for instance in terms of the costs of acquiring each, their contribution to psychological health, and so on, the attitude that is the belief that *P brings with it* a particular standard of evaluation. That standard is truth. (Actually, this is a simplification in order to make the point. I shall return to the complications involved in specifying the constitutive standard of attitudes below. For now, the idea that the constitutive standard of belief is truth will be sufficient.) What does it mean to say that the belief that *P brings with it* the standard of truth? It means that part of *what it is* to be a belief that *P* is to be subject to this standard of evaluation. A mental attitude that we did not think of as liable to this standard would not be one that we thought of as a member of the kind ‘belief’: being a regarding as true that is a member of the kind ‘belief’ means, in part, being liable to the standard of evaluation of truth. Again, this is not to say that we cannot similarly evaluate an imagination in terms of how well it conforms to the standard of truth. Perhaps by some chance what is imagined is in fact true, and so the imagination is adequate, or even exceptional, in regards this standard. This is to say that an imagination is not, simply in virtue of the kind of mental attitude it is, the kind of mental attitude that is liable to such evaluation, whereas a belief is so liable. And this is not to say that we cannot evaluate belief in other terms, such as whether it is costly, or

whether it contributes to psychological health. This is to say that the belief is not, simply in virtue of the kind of thing it is, the kind of thing that is liable to such evaluation.

We can use our sharpened language to express the point. We can say that belief is subject to a *constitutive* standard of correctness involving truth, whereas imagination, while it might be subject to some standards of correctness or other – including perhaps the standard of truth – is not subject to any *constitutive* standard of correctness. The standard of truth expresses a constitutive ideal for belief – a way belief must be, if it is to be a correct instance of the kind of mental attitude it is. In order for the belief that *P* to meet the standards to which it is subject *qua belief that P* it has to be true. That is what it means to say that truth is the constitutive standard of correctness for belief.

How do these observations about belief and imagination help with our original puzzle? Recall, that puzzle was motivated by noticing a difference between attitudes: some attitudes, but not others, are liable to the Wrong Kind of Reasons Problem. In particular, belief, but not imagination, is liable to the Wrong Kind of Reasons Problem. The puzzle is the puzzle of explaining why in general this is so. We just observed a different difference between belief and imagination: belief, but not imagination, is subject to a constitutive standard of correctness. Here, then, is my hypothesis: it is the latter difference that explains the former. That is, the fact that belief is subject to a constitutive standard of correctness is what makes it liable to the Wrong Kind of Reasons Problem, and the fact that imagination is not similarly subject to a constitutive standard of correctness is what makes it not liable to the Wrong Kind of Reasons Problem. Quite generally, my hypothesis is that what explains why some attitudes, but not others, are liable to the Wrong Kind of Reasons Problem is that some attitudes, but not others, are subject to constitutive standards of correctness.

Why would being subject to a constitutive standard of correctness make an attitude liable to the Wrong Kind of Reasons Problem and failure to be subject to such a standard make an attitude not liable to the problem? The answer is simple: the distinction between the right and the wrong kind of reasons is the distinction between reasons that are evidence that an object conforms to a constitutive standard of correctness and reasons that are not such evidence. Thus, in the case of attitudes that lack a constitutive standard of correctness, the distinction between ‘right’ and ‘wrong’ kinds of reason is inapt, since making sense of that distinction requires, first of all, that the attitude in question is subject to a constitutive standard of correctness.

As confirmation of this idea, consider, again, the right and the wrong kind of reasons for belief that *D*, the Democrats will retain control of the Senate in 2014. Recall, the right kind of reasons for believing *D* are facts like:

- (i) Recent polling data suggests that *D*.

- (ii) Some reliable political analysts believe that *D*.
- (iii) The number of vulnerable Democratic seats is insufficient to result in a Republican takeover.

What (i-iii) have in common is that each of them is evidence that a belief that *D* is *true*, i.e. is evidence concerning whether a belief that *D* conforms to a constitutive standard of correctness for belief. My suggestion is that this fact, that each of (i-iii) is evidence that the belief that *D*'s conforms to a standard of correctness constitutive of belief, is what makes (i-iii) the right kind of reasons for the belief that *D*. The wrong kind of reasons for believing *D* are facts like:

- (iv) It is pleasant to believe that *D*.
- (v) Your peer group will like you more if you believe that *D*.
- (vi) Democratic strategists will torture your family unless you believe that *D*.

What (iv-vi) have in common is that each of them is evidence that a belief that *D* is *useful*, i.e. is evidence that a belief that *D* conforms to a standard of correctness for belief we might call the standard of *utility*. But the standard of utility is not a standard that applies to belief *as such*, simply in virtue of the kind of attitude it is. My suggestion then is that this fact, that each of (iv-vi) is evidence that the belief that *D* conforms to a standard of correctness that is *not* constitutive of belief, is what makes (iv-vi) the wrong kind of reasons for the belief that *D*. Importantly, *this is not to say that (iv-vi) are not reasons for belief*. This is to say that they are the *wrong kind of reasons* for belief, because they are not evidence that the belief conforms to a standard of correctness that applies to it simply in virtue of the kind of attitude it is. So that is how the account solves the Wrong Kind of Reasons Problem for belief.

This hypothesis, along with its attendant solution to the Wrong Kind of Reasons Problem, predicts that for attitudes like imagination, which lack a constitutive standard of correctness, we will be unable to distinguish between the right and the wrong kind of reasons for the attitude. And we have already seen that this is so in the case of imagination. The hypothesis also predicts that, wherever we are able to distinguish the right from the wrong kind of reasons for an attitude, there too we will think that the attitude is governed by a constitutive standard of correctness. Is this prediction borne out? I think it is. Obviously I cannot proceed stepwise through all attitudes, showing for each one that is liable to a Wrong Kind of Reasons Problem that it is also subject to a constitutive standard of correctness. But let me provide some inductive evidence for my proposal by illustrating how it deals with a range of attitudes, including one we are already familiar with, one we we know is liable to the Wrong Kind of Reasons Problem: admiration.

According to the proposal on offer, an attitude is liable to the Wrong Kind of Reasons Problem if and only if, and because, it is subject to a constitutive standard of correctness, i.e. a standard of correctness that applies to the attitude simply in virtue of being the kind of attitude it is. So: are all attitudes liable to the Wrong Kind of Reasons Problem subject to such constitutive standards? I think we have good independent reason to say they are. Consider a range of different attitudes you might have toward some person, *N*. You might fear *N*, trust *N*, hate *N*, love *N*, despise *N*, envy *N*, respect *N*, admire *N*... the list continues. What distinguishes one attitude from another? In particular, is there something analogous to 'being true' or 'being an accurate map of the London Underground' that, as a standard for each of these attitudes, helps constitute the attitude as the distinctive attitude it is?

We can begin to answer this question by noticing that each of these attitudes represents its objects as being a certain way. Fear represents its objects as *dangerous*, trust as *trustworthy*, envy as *envious*, admiration as *admirable*, and so on. Part of what it is to be the distinctive attitude each of these attitudes is seems to be for the attitude to represent its object as being the distinctive way each attitude represents its object as being, viz. dangerous, trustworthy, envious, admirable, and so on. Here, then, is my suggestion: this is best thought of in terms of there being a constitutive standard of correctness for each of the attitudes in question. So: the constitutive standard of correctness for the attitude of *fear* is *dangerousness*; for *trust*, *trustworthiness*; for *envy*, *enviousness*; for *admiration*, *admirability*. Putting the point this way means that admiration is distinguished from, say, trust, in part by the former, but not the latter, being subject to a standard of admirability.

This is not to say that, when it comes to admiration, we are always exclusively interested in admirability. Indeed, we regularly wonder whether admiring someone is to our advantage, or would offend them, or would please those around us. Practical advantage, avoiding offense, peer approval, and so on are all what we might think of as *standards of correctness* for admiration: they are all standards the meeting of which sometimes matters to us when it comes to admiration. But the standard of admirability plays a special role for admiration, analogous to the role played by the standard of truth for belief: thinking of an attitude as the attitude of admiration already involves thinking of it as subject to the standard of admirability, just like thinking of something as a belief already involves thinking of it as evaluable in terms of whether or not it is true. Part of what it is for admiration to be the kind of mental attitude it is – or, we might say, what turns admiration from an attitude that merely involves fondly regarding its object into an instance of *admiration* – is for it to be subject to a standard of correctness involving whether or not its object is admirable. Admirability in this sense expresses a constitutive ideal for the attitude of admiration. Similar remarks apply to the other attitudes. What turns mere negative affect into the attitude of envy?

The thought that the attitude is subject to a standard of correctness involving whether or not its object is in fact *enviable*.

Notice that what this means is that, in order to provide a complete solution to the problem, we shall have to have an *independent* account of the properties that comprise the relevant constitutive standards for the attitudes in question. If the right kind of reasons for admiration are those reasons that are evidence that the object of admiration is *admirable* then in order to identify the right kind of reasons for admiration we shall need an account of what *admirability* comprises. Importantly, *the very same thing goes for the attitude of belief*. Earlier, I said that the idea that 'truth' comprised the constitutive standard for belief was a simplification. We are now in a position to see what sort of simplification this was, and why it was merited. In fact, all that we are strictly licensed to say about the attitude of belief is that it is a mental attitude subject to the constitutive standard of *credibility*. The simplification I made was in assuming a substantive normative view according to which *credibility* is a matter of *truth*. This simplification is justified, I think, in virtue of the fact that there is wide – perhaps universal agreement – on this idea, viz. that *credibility* – meriting belief – is a matter of being true. But there could in principle be disagreement on this fact: a bizarro-epistemologist might implausibly suggest, for instance, that *credibility* – what merits belief – is *falsehood* rather than truth.³⁵ We would presumably have no truck with such a view, since we would reject its first-order normative implications to the effect that, e.g., evidence that something is false is the right kind of reason to believe it. The point here is just that such a view is possible: what determines the content of the right kind of reasons for an attitude is in part a substantive normative view about which properties comprise the constitutive standards for the attitude in question. There is not, or at least not much, disagreement in the case of belief over what the properties are that comprise 'credibility': *credibility* is a matter of truth. But when it comes to attitudes like admiration, fear, envy, and so on, such disagreement is not only possible, but actual.

As I said, the disagreement in these cases is a substantive normative disagreement over what the properties are that comprise the constitutive standards for the attitude in question. So, for instance, there can be substantive normative disagreement over whether some property partially comprises the property of 'admirability,' and then, according to my view, there will be resulting disagreement over what the right kind of reasons are to admire someone. I think it is plausible to suppose that we are in fact in such substantive disagreement with a range of 18th- and even 19th-century philosophers and laymen over whether the property of 'chastity' partially comprises the property of admirability, i.e. whether someone's (in particular a woman's) being chaste contributes to her admirability. What this means is that we will, at the level of the reasons, be in substantive disagreement over whether evidence that someone is chaste is *ipso facto* the right kind of reason to admire her. I take it that we think this is false, and that at least some have

thought it is true.³⁶ To my ear, this sounds like exactly the result we want from an account of the right kind of reasons: we want our account to allow for the possibility of disagreement over what the right kind of reasons are to, e.g., admire someone, and to explain this disagreement in terms of a more fundamental substantive normative disagreement over what admirability amounts to.

This sort of disagreement over the right and the wrong kind of reasons shows up not just in cases of admiration, but also in cases of other attitudes, such as the attitude of blame. Blame is an attitude for which there can be reasons, and there is a corresponding distinction between the right and the wrong kind of reasons to blame someone. And there can be substantive normative disagreement over whether some property partially comprises the property of blameworthiness. For instance, we might all agree that the fact that it will lead to excellent outcomes in terms of utility is a good reason to blame someone, i.e. a relatively strong reason to do so. (We might not agree to this, but let's say we do for the sake of argument.) But what (certain sorts of) utilitarians and (certain sorts of) deontologists disagree over is whether such a reason is a reason of the *right kind* to blame someone: utilitarians think of blameworthiness as comprised of the property of promoting overall utility; deontologists think of blameworthiness as comprised of the property of having (say) violated a duty.³⁷ What this means is that utilitarians and deontologists are, at the level of the reasons, in substantive disagreement over whether evidence that blaming someone promotes overall utility is *ipso facto* the right kind of reason to blame her. Deontologists think this is false, and (at least some) utilitarians think it is true. Again, this sounds like the right result: even if deontologists were to grant that promoting overall utility generates a (possibly strong) reason to blame someone, their disagreement with utilitarians (on this point, at least) is over whether such a reason is a reason of the right kind to do so. (And settling whether it is a reason of the right kind will not, of course, settle whether or not we ought to blame the person.)³⁸

So much for the possibility of disagreement about the right and the wrong kind of reasons. Let me try and convince you that my account delivers the intuitively correct results about the right and the wrong kind of reasons for admiration. As we now know, doing so will require assuming some substantive normative claims about what admirability comprises. I shall try and keep these assumptions as uncontroversial as possible. Recall, intuitively, the right kind of reasons for admiring *N* are facts like:

- (i) *N* is courageous.
- (ii) *N* is kind.
- (iii) *N* is generous.

What (i-iii) have in common, I suggest, is that each of them is evidence that admiring *N* is admiring something *admirable*, i.e. is evidence that admiring *N* conforms to a constitutive standard of correctness for admiration.

My idea, then, as in the case of belief, is that this fact, that each of (i-iii) is evidence that admiring *N* conforms to a standard of correctness for admiration is what makes (i-iii) the right kind of reasons to admire *N*. The wrong kind of reasons to admire *N* are facts like:

- (iv) It is pleasant to admire *N*.
- (v) *N*'s benefactor will pay you to admire her.
- (vi) Criminals will torture your family unless you admire *N*.

What (iv-vi) have in common is that each of them is evidence that admiring *N* is in some way *useful* or *expedient*, i.e. is evidence that admiring *N* conforms to a standard of correctness for admiration we might call the standard of *utility*. But the standard of utility is not a standard that applies to admiration *as such*, simply in virtue of the kind of attitude it is. My suggestion then is that this fact, that each of (iv-vi) is evidence that admiring *N* conforms to a standard of correctness that is *not* constitutive of admiration, is what makes (iv-vi) the wrong kind of reasons for admiring *N*. Importantly, again, this is not to say that (iv-vi) are not reasons for admiring *N*. This is to say that they are the *wrong kind of reasons* for admiring *N*, because they do not constitute evidence that the attitude conforms to a standard of correctness that applies to it simply in virtue of the kind of attitude it is. So that is how the account solves the Wrong Kind of Reasons Problem for admiration; and I hope it is clear how it would go for other attitudes too.

4. *Too narrow?*

In developing his solution to the Wrong Kind of Reasons Problem for Attitudes, Schroeder observes that a structurally identical problem arises not just for attitudes, but also for certain activities, such as making a move in chess, tying knots, and setting the table for a White House State Dinner.³⁹ His idea is that, just as there can be the right and the wrong kind of reasons to admire someone, there can be the right and the wrong kind of reasons to make a move in chess. For instance, intuitively, the fact that castling is likely to lead to checkmate is the right kind of reason to castle in a game of chess, whereas the fact that someone will pay you to castle is the wrong kind of reason to castle in a game of chess. Moreover, this difference between the right and the wrong kind of reasons in chess seems structurally the same as the difference between the right and the wrong kind of reasons when it comes to attitudes like belief and admiration. Schroeder uses this observation to motivate his own strategy for solving the problem over its competitors, since he thinks that his strategy, but not his competitors', can account for the difference between the right and the wrong kind of reasons for activities as well as

attitudes. His idea, then, is that insofar as we should prefer a solution to the Wrong Kind of Reasons Problem that solves the problem not just for attitudes but also for activities, we should prefer his solution.⁴⁰ Now, we have already seen why Schroeder's solution will not work: either it cannot make good sense of the idea that there is a shared set of reasons for an attitude or activity, or it cannot make sense of the idea that the shared set of reasons there is for an attitude or activity is equivalent to the set of the right kind of reasons. But I think Schroeder is right that we should prefer a solution to the Wrong Kind of Reasons Problem that solves the problem *wherever* it occurs, not just as it occurs with attitudes.

Notice that our preference for a solution to the problem wherever it occurs gives us an additional reason to reject Hieronymi's account. That is because, even if it could somehow be made to work for the range of attitudes for which the problem arises – including attitudes such as admiration which, as I have already argued, present problems for the account – that account has no hope of solving the problem as it arises for activities. That is because it essential to Hieronymi's solution that the relevant phenomenon occur with *attitudes* for which there is a question the answering of which amounts to forming the attitude. Recall: it was this feature of her view that caused trouble in the case of admiration, since there did not appear to be any such question. But even if we could somehow manage to identify a question the answering of which amounted to, e.g., admiring someone, the account cannot be extended to cover the problem as it arises for activities. That is because of the obvious fact that there is no question the answering of which amounts to, e.g., making a move in chess. In any case, my interest here is not in rehearsing the difficulties with Hieronymi's account; instead, I want to investigate how my own solution fares in accounting for the difference between the right and the wrong kind of reasons as that difference arises with respect to activities.

So: can the account I offered above be extended to solve the problem when it comes to activities? I think it can, and that it is clear how such an extension would go. Some activities, a move in chess, tying a knot, and setting the table for a White House State Dinner, *bring with them* particular standards of evaluation. Let me focus on chess. For a move in chess, that standard is, roughly, advancing checkmate. What does it mean to say that a move in chess brings with it the standard of advancing checkmate? It means that part of *what it is* to be a move in chess – as opposed to merely an instance of moving an oddly shaped piece of material around a checkered board – is to be subject to this standard of evaluation. A moving of an oddly shaped piece around a checkered board that we did not think of as liable to this standard would not be one that we thought of as a member of the kind 'a move in chess'. We can put the point thus: a move in chess is subject to a *constitutive* standard of correctness involving advancing checkmate, whereas simply moving oddly shaped pieces around the board is not. The standard of advancing checkmate expresses a constitutive ideal for a move in chess – a

way a move in chess must be, if it is to be a correct instance of the kind of activity it is.

And now, with these observations in front of us, it should be clear how the resulting account of the right and the wrong kind of reasons for a move in chess will go. Something is the right kind of reason to make a move in chess – castling, say – just in case it is evidence that making that move conforms to the constitutive standard of correctness for moves in chess, viz. advancing checkmate. And something is the wrong kind of reason to make a move in chess otherwise. This is not to say that someone's paying you is not a reason to castle; this is to say that it is the *wrong* kind of reason to castle, because the fact that someone will pay you to castle does not bear on whether castling conforms to the constitutive standard of correctness for moves in chess, namely, advancing checkmate. It should be clear, then, how my account can be extended to handle the case of the Wrong Kind of Reasons for Activities.

5. *Concluding remarks*

So, unlike both Hieronymi and Schroeder's views, my view delivers the extensionally correct results about what the right and the wrong kind of reasons are for attitudes like belief, admiration, and so on. And unlike Hieronymi's view, it can be extended to capture what is plausibly the same distinction as it arises in the case of activities. At this point you might have the following thought: who cares? More specifically, you might think: Why should I care whether or not we have a view that can deliver the extensionally correct results about the right and the wrong kind of reasons for attitudes (and activities)? My response to this is that, insofar as we are interested in providing a unified, systematic account of the nature of reasons, an account that provides a unified, systematic account of an intuitive distinction among reasons is of intrinsic interest. Recall, we began the article by considering two intuitive instances of the distinction between the right and the wrong kind of reasons. As I mentioned, you might have been worried by the idea that there was no unified explanation of what was going on in these cases. In response to this worry, I appealed to explanatory unity and offered a promissory note. We are now in a position to cash that note in: if I am right, then we have before us a systematic, unified explanation of what it is for something to be the right or the wrong kind of reason for an attitude (or activity): what it is for something to be the right kind of reason for an attitude (or activity) is for it to be evidence that the attitude (or activity) conforms to its constitutive standard of correctness.

I argued here that Hieronymi and Schroeder's solutions to the Wrong Kind of Reasons Problem do not work. Let me close by commenting again

on what I think Hieronymi and Schroeder each get right, and on how my view improves on each of their accounts.

Hieronymi says that the right kind of reasons for an attitude bear on a question. According to Hieronymi, the question the right kind of reasons bear on is a question the answering of which amounts to forming the relevant attitude. As we saw, the problem with this idea is that at least some attitudes that are liable to the problem are not ones for which there is a question the answering of which amounts to forming the attitude. But with my account in front of us, we can now make good on Hieronymi's idea that the right kind of reasons bear on a question. On my account, the right kind of reasons to have an attitude are evidence that the attitude in question conforms to a constitutive standard of correctness. We could just as easily put this point by saying that the right kind of reasons bear on the question of whether or not the attitude conforms to a constitutive standard of correctness. But, and here is where my disagreement with Hieronymi lies, it is not the case that answering the question of constitutive correctness with respect to any attitude simply *amounts* to forming the attitude. It may work like this for some attitudes – again, belief is a particularly compelling case – but it need not. As we saw, it does not work like this for the attitude of admiration.

Schroeder holds that the right kind of reasons are shared by necessarily anyone engaged in having an attitude and just because they are so engaged. According to Schroeder, these shared sets of reasons exist either because of *background facts* about the attitude in question (the *background facts strategy*), or because of a standing reason against having false, or for having true, mental representations (the *alethic strategy*). As we saw, the problem with Schroeder's account is that neither of these two strategies can do the work establishing both that (i) there is a shared set of reasons for an attitude and (ii) this shared set of reasons is co-extensive with the set of the right kind of reasons. But with my account in front of us, we can now make good on Schroeder's idea that the right kind of reasons are those that are shared by necessarily anyone engaged in having an attitude and just because they are so engaged. On my account, the right kind of reasons to have an attitude are evidence that the attitude in question conforms to a *constitutive* standard of correctness. Because the standard is partly constitutive of the attitude being the attitude it is, anyone engaged in having the attitude will share those reasons, and just because they are so engaged. Thus it might look like my account goes so far as to vindicate Schroeder's view by showing how it is that there are shared sets of reasons for anyone engaged in having an attitude. But this appearance is misleading. That is because, on Schroeder's view, (constitutive) standards of correctness must be explained in terms of shared sets of reasons, whereas, on my view, the existence of shared sets of reasons is explained by the fact that there are constitutive standards of correctness for certain attitudes. And this, in turn, is explained by the way in which we distinguish attitudes of one kind from attitudes of another. That

means, if I am right, then Schroeder's account might well be *extensionally correct*, but it is *explanatorily* backwards: it may get the (right kind of) reasons right, but it does so for the wrong reasons.⁴¹

Department of Philosophy
University of North Carolina at Chapel Hill

NOTES

¹ Some philosophers – fundamentalists about reasons – think this is the most informative analysis possible of reasons possible. For instance, see Scanlon, 1998; Parfit, 2011. Others – anti-fundamentalists about reasons – think we can give an informative, non-trivial account of what the 'counting in favor of' relation amounts to. For instance, see Kearns and Star, 2009; Broome, 2004; Schroeder, 2007a; Sharadin, 2013a. This debate will not concern me here, since the problem I am interested in faces fundamentalists and anti-fundamentalists alike; but see Sharadin, 2013a, in which I argue for an anti-fundamentalist account of reasons that pairs nicely with the solution I offer to the problem that concerns me here.

² Surprisingly, this question has received little attention. Some notable exceptions include Schroeder, 2007b; Kearns and Star, 2013; Broome, 2008.

³ Hieronymi, 2005, 2013; Schroeder, 2007a; Olson, 2004, 2009; Rabinowicz and Rønnow-Rasmussen, 2004, 2006; Stratton-Lake, 2005; Lang, 2008; Reisner, 2009; Schroeder, 2010; Way, 2012. See also D'Arms and Jacobson, 2000, where D'Arms and Jacobson argue against what they call the 'moralistic fallacy,' i.e. the fallacy of inferring from the fact that feeling an emotion would be wrong or vicious to the fact that it is therefore unfitting. The problem I am concerned with in this article is obviously related to the one D'Arms and Jacobson address, but it is not the same. D'Arms and Jacobson are concerned with a particular subset of the 'wrong' kind of reasons (moral and – sometimes – prudential reasons) for a particular subset of attitudes (the emotions). In this article, I aim to address the problem as it appears across the whole range of attitudes (and activities – see §4) for which the distinction between the 'right' and the 'wrong' kind of reasons arises. And my proposed solution to the problem is different from the one D'Arms and Jacobson propose, appealing as it does to a certain distinctive class of standards for attitudes and activities. More on this in §3. Thanks to an anonymous referee for suggesting I clarify the connection between the problem as it arises in the case of emotions and the problem as it arises more generally, as I address it here.

⁴ Thanks to John Roberts for this way of putting the worry.

⁵ Schroeder (2012) calls these the 'earmarks' of the distinction. Hieronymi (2013) argues that quite generally, 'earmarks' are not always trustworthy as a guide to a distinction that interests us, such that an account of the distinction that cannot capture (all) the earmarks of the distinction fails as an account. I am sympathetic to this idea; but Hieronymi's argument is not germane at present, since I am appealing to the earmarks of the distinction only to motivate the idea that there is *some* puzzling phenomenon present, against the view that there is none – an idea Hieronymi surely agrees with, since she herself is in the business of providing an account of the relevant phenomenon.

⁶ Raz (2009) takes this to be *the* distinguishing mark of the right kind of reasons. See also Schroeder, 2012.

⁷ Hieronymi, 2005, p. 447. See also Hieronymi, 2013.

⁸ But see Boyle, 2011, and Cassam, 2010, for some discussions of the complications involved in this idea.

⁹ Hieronymi, 2005, p. 448.

¹⁰ Thanks to Simon Blackburn for this way of putting the point.

¹¹ The details of how one settles for oneself a question do not matter here. What is important is the idea that, for any question, settling it one way or the other for oneself will simply involve forming beliefs whose contents bear on the question or, perhaps equivalently, assenting to or making judgments whose content bears on the question. For more details on how this might work, see Schaffer, 2007.

¹² It is open to Hieronymi to insist on an intellectualized conception of attitudes such that any attitude for which there can be reasons is an attitude such that making certain judgments amounts to having it. There is some evidence that Hieronymi herself has such a view. See for instance Hieronymi, 2013. But because taking this route involves collapsing the distinction between thought and feeling, this would severely limit the appeal of her account. Anyone who thinks, e.g., both that there can be reasons for desire and that having a desire does not simply amount to making certain judgments or having certain beliefs will be unable to accept Hieronymi's account of the problem. For this reason I assume that, even if it is open to *Hieronymi* to pursue this line, it is a flaw with her solution to the problem since accepting that solution will involve accepting controversial claims very few are willing to accept.

¹³ Thanks to an anonymous referee for suggesting this further problem for Hieronymi's account.

¹⁴ I give this same account of what Schroeder's view is, and what is wrong with it, in Sharadin, 2013b.

¹⁵ Schroeder, 2010, p. 13.

¹⁶ Schroeder, 2010, pp. 35–36.

¹⁷ Schroeder, 2010, p. 37. Two remarks are in order. First, Schroeder's official statement of the view leaves open the possibility that, relative to, say, believing, there can be the right kind of reasons for attitudes other than belief. This possibility will not concern me here, and it is not relevant to the problems I raise with his solution. So I elide this possibility in my statement of Schroeder's principle. Second, Schroeder's account is actually designed to apply more generally, beyond the case of attitudes. Officially, his view is that, relative to an *activity* A, R is the right kind of reason to ϕ iff R is a reason shared by necessarily anyone engaged in A and just because they are so engaged. But, restricting our attention just to his solution as it applies to attitudes, Schroeder's account is as I have it. I return to the question of whether the correct solution to the problem will also apply to activities below, in §4.

¹⁸ Schroeder, 2010, p. 42.

¹⁹ Schroeder, 2010, p. 42.

²⁰ Schroeder's reason for thinking there is a set of reasons shared by necessarily anyone engaged in emulating is presumably that, first, one does not even count as engaged in emulating unless one also aims at emulating those it would not be a bad idea to emulate, and that, second, having an aim of this sort guarantees the presence of certain reasons. See Schroeder, 2007a, ch. 7, esp. p. 135 and following.

²¹ Thanks to an anonymous referee for pressing me to be clearer about this point.

²² Thanks to Derek Baker for this way of putting the point.

²³ Schroeder, 2010, p. 42.

²⁴ Of course, it is not true that in every case admiring someone involves representing them as generous. This is because there can be disparate grounds for finding someone admirable, such as their kindness, courage, or strength. I make the simplifying assumption for the sake of easing exposition. In its expanded form, the argument would begin with the assumption that admiring someone involves representing them as generous, or kind, or courageous, or strong, or... and so on for all the properties the having of which we think is admirable. The argument would then proceed as before: the alethic strategy explains why the fact that someone is not generous, or not kind, or not courageous, or not strong, and so on... is the right kind of reason *against* admiring that person, but it does not explain why the fact that someone *is* generous, or is kind, or is courageous, or is strong, is the right kind of reason to admire them. In what follows, I shall continue to

use the abbreviated form of the argument. Thanks to an anonymous referee for recommending clarity on this point.

²⁵ Thanks to an anonymous referee for this suggestion on behalf of the alethic strategy.

²⁶ See Schroeder, 2007b, pp. 121–124; 2007a esp. chs. 5 and 7; and Schroeder, 2007c.

²⁷ Again, Schroeder's account is meant to apply to reasons for *action* as well as reasons for *attitudes*. But my focus here is solely on the latter.

²⁸ Schroeder, 2007b, p. 123.

²⁹ For reasons of space, I will not argue for this principle here. For some intuitive support in its favor, notice that, if it were false, then there could be an overwhelmingly strong reason to perform an action that is a means to performing an action there is an underwhelmingly weak reason to perform. Intuitively, at least, this is the wrong result.

³⁰ I am grateful to an anonymous referee for this way of putting the problem for the alethic strategy.

³¹ Schroeder, 2010, p. 13.

³² I borrow this way of putting the point from Shah and Velleman, 2005.

³³ Schroeder also makes this point about the ease of generating the wrong kind of reasons for attitudes by way of introducing extraordinary incentives in Schroeder, 2010. According to Schroeder, that is part of the clue that what is going on in these cases has to do with the reasons being idiosyncratic.

³⁴ Thus, a map of the London Underground is what Judith Thomson calls a 'correctness-fixing kind'. For Thomson's account of correctness-fixing kinds, see Thomson, 2008. I disagree with Thomson on the correct diagnosis of the Wrong Kind of Reasons Problem, but I agree with her, to a large degree, on what it takes for something to be liable to the problem, viz. that it must be a member of a correctness-fixing kind.

³⁵ Somewhat, but only somewhat, less implausibly, Richard Rorty infamously floats a view according to which what merits belief is truth, but according to which truth is correctly understood not as matter of correspondence to reality, but instead as a matter of that which one's contemporaries are willing to accept. See Rorty, 1989.

³⁶ Hume seems to be among those that thinks chastity is admirable; he says, at least, that it is praiseworthy. See Hume, 1978, 3.2.12.

³⁷ At least, this is what utilitarians say they think though, as an anonymous reviewer pointed out to me, it may be that such utilitarians are making a conceptual mistake about blame and what they really think is that blameworthiness doesn't matter at all. Instead, what matters is who it is optimal to blame. But for present purposes I'll take them at their word.

³⁸ Thanks to an anonymous referee for taking issue with an earlier version of this example and pressing me to be clearer on this point. For another example of a substantive normative view about the content of the properties that comprise the constitutive standards of an attitude – one with which we might disagree – see Rawls's claims concerning *envy* in Rawls, 1971, ch. 9.

³⁹ Schroeder, 2010, p. 9.

⁴⁰ Schroeder, 2010, p. 10.

⁴¹ Thanks to Simon Blackburn, Finnur Dellsen, Dan Layman, Katherine Nolfi and Geoff Sayre-McCord for valuable discussions and feedback on the issues raised in this article. Thanks also to the Charlotte W. Newcombe Foundation for its generous support.

REFERENCES

- Boyle, M. (2011). "“Making up Your Mind” and the Activity of Reason,' *Philosophers' Imprint* 11(17), 1–24.
- Broome, J. (2004). "Reasons," in J. Wallace, M. Smith, S. Scheffler and P. Pettit (eds.) *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press, pp. 28–55.

- Broome, J. (2008). "Reply to Southwood, Kearns and Star, and Cullity," *Ethics* 119, pp. 96–108.
- Cassam, Q. (2010). "Judging, Believing and Thinking," *Philosophical Issues* 20, pp. 80–95.
- D'Arms, J. and Jacobson, D. (2000). "The Moralistic Fallacy: On the "Appropriateness" of Emotions," *Philosophy and Phenomenological Research* 6, pp. 65–90.
- Hieronymi, P. (2005). "The Wrong Kind of Reason," *Journal of Philosophy* 102(9), pp. 437–57.
- Hieronymi, P. (2013). "The Use of Reasons in Thought (and the Use of Earmarks in Arguments)," *Ethics* 1, pp. 114–27.
- Hume, D. (1978). *A Treatise of Human Nature*, 2nd edn., L.A. Selby-Bigge and P.H. Nidditch, (eds.) Oxford: Oxford University Press.
- Kearns, S. and Star, D. (2009). "Reasons as Evidence," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics 4*. Oxford: Oxford University Press, pp. 215–42.
- Kearns, S. and Star, D. (2013). "Weighing Reasons," *Journal of Moral Philosophy* 10, pp. 70–86.
- Lang, G. (2008). "The Right Kind of Solution to the Wrong Kind of Reason Problem," *Utilitas* 20(4), pp. 472–89.
- Olson, J. (2004). "Buck-Passing and the Wrong Kind of Reasons," *Philosophical Quarterly* 54(215), pp. 295–300.
- Olson, J. (2009). "The Wrong Kind of Solution to the Wrong Kind of Reason Problem," *Utilitas* 21(2), pp. 225–32.
- Parfit, D. (2011). *On What Matters*, Volume II. Oxford: Oxford University Press.
- Rabinowicz, W. and Rønnow-Rasmussen, T. (2004). "The Strike of the Demon: On Fitting Pro-Attitudes and Value," *Ethics* 114(3), pp. 391–423.
- Rabinowicz, W. and Rønnow-Rasmussen, T. (2006). "Buck-Passing and the Right Kind of Reasons," *Philosophical Quarterly* 56(222), pp. 114–20.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press.
- Raz, J. (2009). "Reasons: Practical and Adaptive," in D. Sobel and S. Wall (eds.) *Reasons for Action*. Cambridge: Cambridge University Press.
- Reisner, A. (2009). "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem," *Philosophical Studies* 145, pp. 257–72.
- Rorty, R. (1989). *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schaffer, J. (2007). "Knowing the Answer," *Philosophy and Phenomenological Research* 75(2), pp. 383–403.
- Schroeder, M. (2007a). *Slaves of the Passions*. Oxford: Oxford University Press.
- Schroeder, M. (2007b). "Weighting for a Plausible Humean Theory of Reasons," *Noûs* 41, pp. 110–32.
- Schroeder, M. (2007c). "The Negative Reason Existential Fallacy," unpublished Ms.
- Schroeder, M. (2010). "Value and the Right Kind of Reason," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics 5*. Oxford: Oxford University Press, pp. 25–55.
- Schroeder, M. (2012). "The Ubiquity of State-Given Reasons," *Ethics* 122(2), pp. 457–88.
- Shah, N. and Velleman, D. (2005). "Doxastic Deliberation," *Philosophical Review* 114(4), pp. 497–534.
- Sharadin, N. (2013a). 'Structuralism About Reasons,' unpublished Ms.
- Sharadin, N. (2013b). "Schroeder on the Wrong Kind of Reasons Problem for Attitudes." *Journal of Ethics and Social Philosophy* 7, pp. 1–8.
- Stratton-Lake, P. (2005). "How to Deal with Evil Demons: Comment on Rabinowicz and Rønnow-Rasmussen," *Ethics* 115(4), pp. 788–98.
- Thomson, J.J. (2008). *Normativity*. Chicago, IL: Open Court.
- Way, J. (2012). "Transmission and the Wrong Kind of Reason," *Ethics* 122(3), pp. 489–515.