

Creditworthiness and Matching Principles

Jonathan Way

University of Southampton

Forthcoming in *Oxford Studies in Normative Ethics*

It's one thing to do the right thing. It's another to do the right thing in a way that reflects well on you. Moral philosophy's most famous illustration of this point is Kant's shopkeeper, who gives his customers the right change but only because this is good for business (Kant 1785 [1997]: 11). When you do the right thing in a way that reflects well on you, your action has *moral worth*, or is *praiseworthy* or, as I shall say, is *creditworthy*.

My question is what makes for creditworthiness. In broad terms, the lesson of Kant's shopkeeper is that creditworthiness requires a certain kind of explanation of *why* you did the right thing. A natural and familiar way to put this is to say that creditworthiness is a matter of doing the right thing for the right reasons. And a simple way to cash out this thought is to say that you are creditworthy just when the reasons for which you act – your motivating reasons – are good reasons to act – normative reasons. This paper focuses on a problem for this simple account. The problem is that, just as it can be an accident that what you did was right, it can be an accident that your motivating reasons were normative reasons. Since this kind of accidentality precludes creditworthiness, the simple account must be rejected. I consider whether the problem can be solved by views which require that you know that the reasons for which you act are normative reasons and argue that it cannot.

Instead, I argue, we should accept a view on which creditworthiness requires that you act from the right principles, as well as the right reasons.

One preliminary point: it's natural to think that creditworthiness comes in different flavours – a response might be *morally* worthy, *prudentially* worthy, *epistemically* worthy, and so on. My discussion will proceed on this assumption, along with the natural corollary that rightness and reasons also come in different flavours, in a parallel way. I am interested in the prospects of accounts that apply to all of these kinds of creditworthiness.

1. The Matching Reasons Account of Creditworthiness

When you act intentionally, there are, at least typically, reasons for which you act. These are motivating reasons. When an action is right, there are reasons to perform that action. These are normative reasons. A natural thought is thus that creditworthiness is just a matter of *matching reasons* – of having normative reasons as your motivating reasons. That is:

(Matching Reasons) You are creditworthy for ϕ -ing iff the (motivating) reasons for which you ϕ are (normative) reasons to ϕ .¹

Three clarificatory points are worth making. First, to allow for different flavours of creditworthiness, the view must be qualified. For instance, moral creditworthiness requires that your motivating reasons are moral reasons, prudential creditworthiness requires that your motivating reasons are prudential reasons, and so on. Second,

¹ For versions of this view, see Arpaly 2002, Arpaly and Schroeder 2014, and Markovits 2010. However, the views of these authors are more sophisticated than the view stated here.

depending on one's view of the ontology of normative and motivating reasons, it might be better to formulate the view so as to require that your motivating reasons *represent* normative reasons. This issue won't matter here, and so I'll stick to the simpler formulation.² Third, Matching Reasons is intended as an account of *pro tanto* creditworthiness – creditworthiness in some respect or to some extent. It is therefore not an objection to Matching Reasons that it allows you to be creditworthy even when you act for reasons which are outweighed. Such acts can still reflect well on you to some extent, even if they do not reflect well on you overall.

Matching Reasons is a simple and elegant way to cash out the natural thought that creditworthiness is a matter of acting for the right reasons. It also offers the right verdict on the shopkeeper case, and other central cases in which a person fails to be creditworthy despite doing the right thing. The shopkeeper's reason for giving the right change is that doing so will be good for business. Since this is not a moral reason to give the right change, it follows from Matching Reasons that he is not morally creditworthy. (He might be prudentially creditworthy).

Matching Reasons has several further attractions too. Some of these stem from its relative undemandingness. All that Matching Reasons requires for creditworthiness is that a person's motivating reasons are normative reasons. It doesn't require that the person know or even believe that their motivating reasons are normative reasons. It thus allows for *unreflective creditworthiness* – creditworthiness without the belief that your motivating reason is a normative reason. Perhaps the best examples are cases of what Nomy Arpaly (2002: 51-3) calls *fast action* – for instance, making shots in a tennis match, and contributions to conversation. Acts of this kind can be morally creditworthy – think of a comforting remark – and they can plausibly be creditworthy

² For discussion, see, e.g. Dancy 2000, Alvarez 2010, Mantel Forthcoming a.

even if done unreflectively. For the same reason, the view also allows for *creditworthy akrasia* – creditworthiness in cases where the person thinks that what they are doing is entirely wrong. To use the standard example, Huck Finn might be creditworthy for helping Jim escape, even if he thought that doing so was entirely wrong, so long as he was moved by reasons which favour helping Jim (Arpaly 2002: 75-8).

It is also worth emphasizing the ease with which Matching Reasons applies to the different flavours of creditworthiness. Note, for example, that the view is not limited to creditworthy *action* – it extends straightforwardly to creditworthy belief and emotions. For example, the view can say that a belief is epistemically creditworthy iff held for epistemic reasons. Indeed, Schroeder (2015) claims that it is an advantage of his account of *knowledge*, on which knowledge is belief held for reasons which are (sufficient, objective and subjective) normative reasons, that it implies that knowledge is creditworthy belief, and thus valuable, given a view like Matching Reasons.³ Since it seems a reasonable presumption that accounts of the different flavours of creditworthiness should be unified, this is a further important attraction of Matching Reasons.

Matching Reasons thus has significant attractions. However, as it stands it is too simple. In the next section, I explain a central problem with it.

2. Matching Reasons Is Not Enough: The Problem of Accidentality

As noted, Matching Reasons is relatively undemanding. All that it requires for creditworthiness is that your motivating reasons match normative reasons. It puts no

³ Markovits (2010: 214) agrees that a view like Matching Reasons gives the right account of epistemic creditworthiness. Compare also Arpaly and Schroeder's unified account of thinking and acting for reasons (2014: ch.3).

constraints on *how* this match comes about. As a result Matching Reasons allows for creditworthiness in cases where it is just an *accident* that your motivating reasons match normative reasons. But it is very plausible that this kind of accidentality precludes creditworthiness. This is perhaps easiest to bring out with an epistemic example:

Sarah knows that there's a spiderweb in the corner and that if there's a spiderweb in the corner, there's a spider nearby. Sarah reasons from this knowledge to the belief that there's a spider nearby. She does so by following the *arachnaphobe's rule*: from any premise concerning spiders, infer that there's a spider nearby.

Sarah believes that there's a spider nearby for the reason that there's a spiderweb in the corner and that if there's a spiderweb in the corner, there's a spider nearby. This is a normative reason – indeed, a conclusive reason – to believe there's a spider nearby. But Sarah does not seem creditworthy. Plausibly, this is because it's too much of an accident that the premises from which she infers provide reason to believe there is a spider nearby. Given the inference rule she applies, she could easily come to the same conclusion from premises which do not provide reason for that conclusion.⁴

We get structurally similar cases in the moral domain:

Billy is wondering who to ask to sing first at his office karaoke party. He knows that Joe is shy, and so decides not to ask him. He makes this decision by following the *shyness rule*: if someone is shy, ignore them.

⁴ For similar cases, used for a variety of purposes, see Turri 2010, Lord and Sylvan ms, and Millar 1991: 57. For similar cases in the moral domain, see Lord forthcoming, Mantel forthcoming b, Stratton-Lake 2000: 65, Markovits 2010: 227.

Billy decides not to ask Joe to sing first for the reason that Joe is shy. This is plausibly a normative reason not to ask Joe to sing first. But again Billy doesn't seem creditworthy. And a similar diagnosis seems to apply. Though the fact that someone is shy is *sometimes* a reason not to involve them in things, it is very often not. Thus given the rule Billy followed, he could easily have decided not to involve Joe in something on the basis of a consideration which did not support so deciding.⁵

Can proponents of Matching Reasons resist these counter-examples? Here are two ways in which they might try to do so.

First, it might be claimed that the cases trade on incomplete specifications of motivating reasons. Strictly speaking, Sarah's reason is not just that there's a spider web in the corner and that if this is so, there's a spider nearby. Her reason also includes the further premise that since these facts concern spiders, there's a spider nearby. Since this further premise is false, Sarah's reason does not match a normative reason. Similarly, Billy's reason is not, strictly speaking, just that Joe is shy. His reason also includes the claim that whenever someone is shy, you should ignore them, or something to that effect. Since this further claim is false, Billy's reason does not match a normative reason.

This response is unsatisfactory. It's a familiar point that we need to distinguish between our premises and the principles or rules we apply in theoretical reasoning. Insisting that our premises must always include a statement of our inference rules leads to an immediate regress (Carroll 1895). And given that this distinction is

⁵ The kind of accidentality involved in these examples is not quite the same as in the case of Kant's shopkeeper. The shopkeeper acts for a consideration which is not a moral reason but which nonetheless leads him to do what there is moral reason to do. Sarah and Joe respond to considerations which are reasons of the relevant kinds but it's an accident that they responded to considerations which are reasons.

required to understand theoretical reasoning, it is hard to see why it should not also apply in practical reasoning.

Second, it might be claimed that the cases trade on incomplete specifications of normative reasons. Strictly speaking, the normative reason for Sarah to believe that there's a spider nearby is not simply that there's a spider web in the corner and that if this is so, there's a spider nearby. The reason also includes the fact that these considerations provide evidence that there's a spider nearby, or something to that effect. Similarly, the normative reason for Billy to decide against asking Joe is not simply the fact that Joe is shy. The reason also includes the fact that, since Joe is shy, singing first would make him uncomfortable. Thus since neither Sarah nor Billy's motivating reasons match complete specifications of normative reasons, neither is creditworthy.

While some might be willing to accept this objection, it comes with considerable cost. That's because it is very plausible that Sarah and Billy's motivating reasons here are normative reasons – it is entirely natural to say, for instance, that the fact that Joe is shy is a reason to decide not to ask him to sing first. Nor does accepting this claim require us to claim that the fact that Joe is shy is the *only* reason to decide not to ask him to sing first, or to deny that the fact that singing first would make Joe uncomfortable is also relevant.⁶

There are also theoretical reasons to allow considerations such as the fact that Joe is shy to count as reasons. In general, it must be possible to act for a reason *r* in a creditworthy way without being aware of, or responding to, all the facts which help explain why *r* is a reason. After all, the full explanation of why some fact is a reason

⁶ That Joe's shyness is sometimes but not always a reason is an example of what Dancy (1993, 2004) calls the 'holism of reasons'. Some philosophers (e.g. Crisp (2000), Raz (2000: ch.10)) have been led to deny holism in order to resist Dancy's case for moral particularism. However, as several authors have shown, we can accept holism while rejecting particularism (McKeever and Ridge 2006: ch.2, Horta 2012: ch.6, Väyrynen 2006).

may well appeal to some fundamental moral principle(s). But you need not be aware of the fundamental moral principle(s) in order to be creditworthy (cf. Star 2015: ch.1). If creditworthiness is possible then, we need a distinction between reasons and considerations which help to explain reasons. But that distinction is all that is needed to generate cases like the arachnaphobe and karaoke cases.⁷

I thus take it that these lines of resistance fail and that the counter-examples do undermine Matching Reasons. However, we also saw that the view has significant attractions. We should therefore consider how it might be revised so as to solve the problem.

3. Normative Knowledge

Famously, Kant thought that the shopkeeper's act lacked moral worth because it was not done *from duty*. Though the shopkeeper's act was right, it was not done *because* it was right. It is natural to move from this to the idea that moral worth requires that you act from *knowledge* that what you are doing is right.⁸

A similar line of thought applies in our cases. Sarah and Billy ϕ in response to considerations which are reasons to ϕ . But they do not ϕ because these considerations are reasons to ϕ . Sarah makes her inference because her premises concerns spiders, not because they provide reason to believe the conclusion. Billy responds to the fact that Joe is shy by deciding not to ask him to sing first because he always ignores shy people, not because this fact supports his decision. What this suggests is that

⁷ However, drawing this distinction does not require us to deny that all such explanatory factors are themselves reasons.

⁸ For recent defence, see Sliwa 2016.

creditworthiness requires that we respond to reasons *because* they are reasons, or *as* reasons; it is not enough that we respond to considerations which are *in fact* reasons.⁹

What we need to understand then, is what it is to respond to a consideration *because* it's a reason to ϕ .¹⁰ And just as it's natural to take doing what's right because it's right to involve doing what's right because you *know* it's right, it's also natural to take responding to a consideration because it's a reason to involve responding to that consideration because you *know* it's a reason. This leads to the following amendment to Matching Reasons:

(Matching Reasons_{Kantian}) When ϕ -ing for reason r , you are creditworthy for ϕ -ing iff (i) r is a reason to ϕ and (ii) you ϕ because you know that r is a reason to ϕ .¹¹

This view is in one way moderate. A stronger Kantian view says that creditworthiness requires that the only reason for which you ϕ is that ϕ -ing is right. By contrast, Matching Reasons_{Kantian} allows that you can be creditworthy in responding to other considerations. I take this to be an advantage. It is very plausible that good people do not *only* respond to rightness as such – they also respond to the suffering of others, promises made, rights ignored, and so on.¹² Matching Reasons_{Kantian} allows this.

⁹ Does this mean that creditworthiness is not to be equated with acting for the right reasons? I think this latter phrase can be taken to mean 'acting for what are in fact the right reasons'. So understood, the equation fails. But it can also be taken to mean 'acting for the right reasons, because they are the right reasons'. Understood in this latter way, the equation can be maintained.

¹⁰ Several authors allow normative causation – they hold that normative properties can be causally efficacious or relevant (Arpaly 2006, Arpaly and Schroeder 2014, Lord and Sylvan ms, Sturgeon 1986, Wedgwood 2006). However, these authors at least typically still require a further account of how this is possible.

¹¹ Weaker versions of this view would only require *belief* or *justified belief* in (i). However, such views would not avoid the problems to be discussed.

¹² Cf. Smith's (1994) discussion of 'moral fetishism' – though Smith's claims go beyond the point in the text.

Still, in allowing this, Matching Reasons_{Kantian} raises the question of what role the normative knowledge is supposed to play. It is not, presumably, supposed to play the same role as knowledge that, for example, this person is suffering. It would be an odd kind of overdetermination to hold that to be creditworthy for responding to someone's suffering one must have at least two reasons for so responding – that the person is suffering and also that this is a reason to help them.

Here is a suggestion for how proponents of Matching Reasons_{Kantian} might think of the role of the normative knowledge.¹³ Suppose that, in general, φ -ing for the (motivating) reason r involves being caused, in the right kind of way, to φ by the belief that r is a reason to φ .¹⁴ Given this view, then when you φ from knowledge that r is a reason to φ , you will thereby be φ -ing for a motivating reason which is also a normative reason. Furthermore, insofar as it is plausible that when your knowledge that p causes you to φ (in the right kind of way), then you φ because p , it seems to follow that when you φ from knowledge that r is a reason to φ , you φ because r is a reason to φ .

This suggestion gives normative knowledge an intelligible role in a view which avoids the difficulties just noted. However, I doubt that the resulting view is satisfactory.

The crucial question is whether being caused to φ by your knowledge that p is a reason to φ , in the way just outlined, is compatible with following the kind of rules that give rise to the problems in the arachnaphobe and karaoke cases. If it is, then it seems clear that the view does not avoid the counter-examples to Matching Reasons.

¹³ It might be said that the knowledge plays an *enabling* or *background* role. This is fine so far as it goes but it doesn't go very far. Compare: in a weak sense, the fact that you exist is an enabling condition of your creditworthy actions. I take it proponents of Matching Reasons_{Kantian} want knowledge of what's right to play a more substantial role than this.

¹⁴ The view that φ -ing for a reason involves such beliefs is a common one. See, e.g. Alvarez 2010: ch.3, Gregory forthcoming, Raz 1999: ch.2. Scanlon 1998: ch.1

For example, if Sarah reasons to the conclusion that there's a spider nearby by applying the arachnaphobe's rule, then her belief does not reflect well on her. This seems true even if knowledge that her premises provide reason for her conclusion plays the role just outlined.¹⁵ Sarah fails to be creditworthy because she follows the arachnaphobe's rule. Insofar as the view outlined is compatible with her following this rule, it does not solve the problem.

So the proponent of this account should deny that causation by normative knowledge is compatible with following rules like the arachnaphobe's rule. However, it is not clear how this can be denied. One idea might be that if you are caused to believe q by knowing that p is a reason to believe q , then the inference rule you apply is 'from the premise that p is a reason to believe q , infer q '. However, this seems to amount to the view that being caused to believe q by knowing that p is a reason to believe q , in the above way, is incompatible with your applying other inference rules. It thus implies that you cannot be creditworthy by applying the inference rule of modus ponens, or inductive rules, and so forth. This seems revisionary.¹⁶

The view just sketched thus seems unsatisfactory. One might wonder if there are other ways of understanding the role of normative knowledge which avoids this problem. I don't have space to consider this in detail. But I do think there are grounds for doubt. For the source of the problem in the arachnaphobe and karaoke cases are the rules being applied. Insofar as normative knowledge is compatible with applying such rules, we should be skeptical that adding normative knowledge can help. And

¹⁵ If we accept that we can acquire normative knowledge by testimony, then we can make this clearer by supposing that Sarah acquires her knowledge that her premises are reasons by testimony, rather than any understanding of why her premises are reasons.

¹⁶ Another idea might be that the role of normative knowledge sketched here is compatible with following *good* rules (e.g. modus ponens) but not bad ones, like the arachnaphobe's rule. But again it is hard to see how to defend this claim. For example, even if we assumed that knowing that p is a reason to ϕ required some understanding of *why* p is a reason to ϕ , and that this understanding implied that you had internalized a good rule of inference – and note, of course, that these are very strong assumptions – this still doesn't rule out that you might also be capable of following bad rules of inference as well.

insofar as normative knowledge is not compatible with following such rules, we require a plausible story about why not, which does not overgeneralize to rule out following the kinds of rules which plausibly do make for creditworthiness.

There are also further difficulties with Matching Reasons_{Kantian}. In particular, the requirement of normative knowledge undermines some of the attractions of the Matching Reasons. First, it rules out the possibility of unreflective and akratic creditworthiness.¹⁷ Second, it does not generalize well. In particular, it leads to a regress when applied to creditworthy belief. On the view under consideration, you are creditworthy for ϕ -ing only if you know that r is a reason to ϕ . But it's highly plausible that you know that p only if your belief that p is creditworthy – that is, it reflects well on you, epistemically. So by Matching Reasons_{Kantian}, your knowledge that r is a reason to ϕ implies a further piece of knowledge – that there's reason to believe that r is a reason to ϕ . This knowledge will itself imply a creditworthy belief, which in turn will require a further piece of normative knowledge. And so on.

This second point suggests that there must be an account of creditworthy belief which does not require normative knowledge. But insofar as it is reasonable to expect a unified account of the different flavours of creditworthiness, that creates a presumption against requiring normative knowledge in other cases too. So we should be suspicious of Matching Reasons_{Kantian}.

4. Matching Principles

I will now sketch an alternative way to develop Matching Reasons, which might be seen as a different account of what it is to respond to a consideration because it's a

¹⁷ More precisely, it rules out the possibility of cases of creditworthy akrasia where you think there is no reason for what you are doing.

reason. I'll start by offering an account of creditworthy inferential belief. We'll then see how far this account can be generalized.

Inferential belief has two ingredients: the reasons for which you believe and the inference rule or pattern which you follow. The lesson of the arachnaphobe case is that the latter matters for creditworthiness. We can see this by contrasting that case with someone who infers that there's a spider nearby from her knowledge of the same premises by applying the inference rule of modus ponens.¹⁸ The latter is a clear case of epistemically creditworthy belief.

So creditworthy inferential belief requires that the inference rule you apply has some significant property. What is that property? Why does applying modus ponens make for creditworthiness and applying the arachnaphobe's rule does not?

The simple initial answer is that modus ponens is a good or correct rule of inference, whereas the arachnaphobe's rule is not. As it stands though, this may seem unsatisfying. What makes for a good rule of inference? And why does adding a good inference rule to good reasons make for creditworthiness?

I think we can make progress with these questions by observing the parallel between the following points. First, the fact that there's a spider-web in the corner and that if there's a spider-web in the corner, there's a spider nearby is plausibly a reason to believe that there's a spider nearby *because* (i) this fact is of the form p , and if p , q and (ii) in general, if it's a fact that p and if p , q , then this fact is a reason to believe q . Second, when you infer by modus ponens, you are applying a rule that say, roughly, that it's good reasoning to move from believing propositions of the form p and if p , q to believing q .

¹⁸ That is: the rule of reasoning which says, roughly, that it's good reasoning to move from believing of the form p and if p , q to believing q . I don't mean the logical rule which is more often called 'modus ponens'. For a classic discussion of the distinction between logic and reasoning, see Harman 1986: chs 1 and 2.

So if you infer by modus ponens there is a match between the explanation of why your premises are reasons for your conclusion and the inference rule which takes you from your premises to the conclusion. More precisely, there is a match between the general principle which the explanation appeals to and the general rule you apply. The former says that all facts of a certain form are reasons for a certain kind of belief. The latter says that it's good reasoning to move from beliefs of the same form to that same kind of belief.

This suggests that we can restate the hypothesis that creditworthy inferential belief requires both normative reasons and a correct inference rule as follows:

Your inferential belief in p is creditworthy when held for reason r iff (i) r is a reason to believe p and (ii) the inference rule you apply matches a principle which explains why r is a reason to believe p .

I submit that this is a highly plausible hypothesis about creditworthy inferential belief. Note first that it sorts our cases in the right way. Someone who infers by modus ponens applies an inference rule which matches an explanation of why her premises provide reason to believe her conclusion. Someone who infers by the arachnaphobe's rule does not: the principle that premises about spiders are generally reasons to think there's a spider nearby is false and so cannot explain why the premises in this example provide reason to think there's a spider nearby.

More generally, the hypothesis seems to rule out the kind of accidentality that precludes creditworthiness. You are not creditworthy for a response if the way in which that response was motivated could quite easily have led you to do the wrong thing. But when an inferential belief is held for good reasons and from a good

inference, you could not quite easily have been led to a belief for which there was no reason. Since good inference rules match principles which explain why your premises are reasons, good inference rules can be relied on to lead you to form beliefs on the basis of normative reasons.

However, even if this hypothesis succeeds as an account of creditworthy inferential belief, it might seem limited in its application, insofar as it appeals to the notion of an inference rule. In a natural sense, much that we believe and do is not the product of inference. But I think there are prospects for generalizing the account.

It's a familiar and plausible hypothesis that when a fact is a reason for you to do something, that's because of some more general principle that facts of that sort are reasons to make responses of that sort. For instance, the fact that you promised to referee a paper might be a reason to do so because, in general, if you promised to ϕ , the fact that you promised to ϕ is a reason to ϕ . The fact that it looks to you as if the sun is out is a reason to think the sun is out because, in general, if it looks to you as if p , the fact that it looks to you as if p is a reason to believe p . Call principles of this sort *normative principles*.

It's also a familiar and plausible idea that acting and believing for (motivating) reasons manifests more general dispositions. For instance, when you referee a paper because you promised to do so, that might manifest a general disposition to ϕ when you promised to ϕ . When you believe that the sun is out because it looks to you as if the sun is out, that might manifest a general disposition to believe p when it looks to you as if p . And so on.

We might think of these dispositions as, in a broad sense, principles – we might call them *motivating principles*. Insofar as you are disposed to believe on the basis of perception, you follow the principle: if it looks to you as if p , believe p .

Insofar as you are disposed to act on the basis of your promises, you follow the principle: if you promised to ϕ , then ϕ .¹⁹ This allows us to generalize our hypothesis. Creditworthiness requires that your motivating reasons match normative reasons *and* that your motivating principles match normative principles:

(Matching Principles) Your ϕ -ing for reason r is creditworthy iff (i) r is a reason to ϕ and (ii) the principle from which you ϕ matches a principle which explains why r is a reason to ϕ .

It might be worth noting that those who would rather use the term ‘principle’ in a narrower way can accept the substance of this account. The core presupposition of the account is simply that both normative and motivating reasons ordinarily reflect something more general; the core idea is then that creditworthiness requires a match of these generalities.

I have already explained how Matching Principles gives the right verdict in the arachnaphobe case. It also gives the right verdict in the karaoke case. The principle from which Billy acts says to ignore shy people. But the fact that Joe is shy is not a moral reason because, in general, when someone is shy, that fact is a reason to ignore them; that is not true. It’s a reason because, in this case, Joe’s shyness would make it uncomfortable for him to sing first and, in general, when doing something would make someone uncomfortable, that fact is a reason not to have them do it. Since Billy’s action doesn’t manifest a disposition matching this principle, Matching Principles implies that he isn’t creditworthy.

¹⁹ This is probably too simple. See, e.g. Railton 2005 and Broome 2013: ch.13 for more sophisticated dispositional accounts of principle-following.

3. Objections and Replies

The view I have sketched raises many questions and requires much further development and elaboration. In this section, I raise some worries about it and indicate some lines of response. What I say will inevitably be programmatic and inconclusive. My aim is just to suggest that none of these worries are immediately decisive.

1) A first objection is that the view is overkill:

The problem in the karaoke and arachnaphobe cases is just that the motivating principles in these cases do not reliably correlate with normative reasons. Thus we can handle these cases by saying that creditworthiness requires your motivating principles to *track* normative reasons. There is no need for your principles also to explain normative reasons. Furthermore, this additional claim makes the view very *demanding*. After all, the question of what makes right acts right is the subject of considerable dispute in normative ethics. The principle which explains why some fact is a reason might be very complicated. Even if it is not, it might be obscure what the principle is. It is surely too much to expect ordinary good people to follow principles which match the true normative theory.

We have here both an objection to the Matching Principles view and a suggestion for an alternative. To test this alternative, we must consider cases in which someone acts

from a reliable but non-explanatory principle. However, it's not as easy as might be thought to get such cases in clear view. Let's consider some putative examples.

First, consider someone who acts from a principle of promise-keeping – at least other things equal, this person is disposed to ϕ when they have promised to ϕ , for that reason. You might think this person's principle is reliable – other things equal, promises provide reasons – but non-explanatory. For you might think that what explains why promises provide reasons will be some deeper fact – for instance, that promise-keeping typically respects humanity.

This example is not convincing. We should distinguish between fundamental and derivative explanations. The general principle that promises are to be kept can explain why some particular promise is to be kept even if there is some further explanation of why promises are generally to be kept. Matching Principles says that one's principles must match an explanation of one's reasons. It does not say that one's principles must match a fundamental explanation of one's reasons. This point should also help to assuage the worry that the view is too demanding (cf. Markovits 2010: 228ff).

Second, consider cases of testimony. Suppose that someone follows the principle of ϕ -ing for reason r when their guru tells them to ϕ for reason r . This principle might be very reliable, because the guru might be very reliable at identifying reasons. But it is not plausibly explanatory, even derivatively. For instance, it's not plausible that the guru's testimony is what makes it the case that the person's promises are to be kept.

One worry about this case is that it's not clear that it's possible to follow this principle. We might think that when the guru tells the person to ϕ for reason r , the

person then acts not just for the reason r , but also because their guru told them to ϕ . Let's grant though that this case is possible. I still doubt that it supports the objection.

Several philosophers are skeptical that you can be creditworthy if you act solely on the basis of reliable normative testimony. That is, they doubt that you can be creditworthy for ϕ -ing if you ϕ just because a reliable source told you that you ought to ϕ (Hills 2009, McGrath 2011). Insofar as we share these doubts, I think we should also be skeptical about whether acting from the above principle can be creditworthy. For instance, suppose your guru tells you that you ought to give money to a certain charity for the reason that doing so will save lives. You do not generally follow a principle of saving lives, but in this case you do as your guru tells you and act for that reason. If you wouldn't be creditworthy for acting from simple testimony that you ought to give money to that charity, it's hard to see how you could be creditworthy here. This case thus provides some support for taking creditworthiness to require something more than just reliability.²⁰

Third, consider people who seem creditworthy despite acting on principles which are false, and so non-explanatory. For example, consider someone who acts on the act-utilitarian principle – and suppose that principle is false. While we might not approve of all this person's actions, we will surely be impressed by some of them.

One point to make here is that versions of this challenge arise for many views. For instance, Arpaly and Schroeder (2014) hold that creditworthiness requires that your reasons match normative reasons *correctly conceptualized* – that is, conceptualized as they are by the correct normative theory. This view thus faces a challenge about what to say about people who follow incorrect theories. Such cases

²⁰ Of course, some philosophers allow that you can be creditworthy when acting on normative testimony. However, such philosophers also often hold that normative testimony can help explain why an action is right (e.g. Markovits 2012: 306-7). If so, the cases in question here do not threaten Matching Principles.

also raise questions for Kantian views, on which creditworthiness requires knowledge that what you are doing is right. For – if utilitarianism is false – it is puzzling how someone whose actions are based on a conviction that utilitarianism is true could know that particular actions are right (cf. Star 2008).

How might the challenge be answered? The natural move is to appeal to a notion of approximation: creditworthiness requires that the principles from which you act *approximate* normative principles. Intuitively, someone who acts from utilitarian convictions comes closer to normative principles than someone who acts on egoist convictions – even though both can lead to right actions. This seems a plausible explanation of why the former person seems more creditworthy than the latter for their right actions. Needless to say, more would need to be said about this notion of approximation. But it seems like a start.

I thus doubt that any of these three cases gives us good reason to think that creditworthiness requires principles which only reliably track reasons. While not conclusive, the testimony case gives us some reason to think that reliable tracking is not enough.²¹

2) A second objection says that the view is too demanding in another way:

Matching Principles says that you're creditworthy for ϕ -ing only if your ϕ -ing manifests some general disposition. However, creditworthy actions need not manifest general dispositions; it is possible to be creditworthy even when acting out of character. Imagine a normally stingy person who gives a

²¹ Another kind of case which might test this objection involves principles which are too disjunctive to be explanatory. Consider, for instance, the principle of ϕ -ing when ϕ -ing would either keep a promise or help someone. However, it is difficult to distinguish someone who acts on this principle from someone who has two principles – ϕ -ing when ϕ -ing would keep a promise and ϕ -ing when it would help someone. So it is hard to get a clear view on this kind of case.

homeless person a sandwich, or a normally cowardly person who risks their life jumping onto a train track to rescue a child (Hurka 2006). These acts may be creditworthy even if they don't manifest any more general dispositions. For example, the cowardly person is creditworthy if they are responding to the danger the child is in, regardless of their general dispositions.

I agree that it is possible to be creditworthy when acting out of character. But Matching Principles does not preclude this possibility. This is for two reasons.

First, one's character depends on one's *settled* dispositions. But Matching Principles does not say that creditworthiness requires acting from a settled disposition. It is plausible that our dispositions can be *fleeting*. To see this, note that it is plausible that we can have fleeting final desires. Indeed, the cases above arguably presuppose this – e.g. if the coward's act is not motivated by a final desire to save the child, it does not seem creditworthy. But desires are widely and plausibly held to involve dispositions. So one way in which Matching Principles allows for action to be creditworthy when out of character is for it to manifest a fleeting disposition.²²

Second, character is not just a matter of settled dispositions; it also depends on the strength of your dispositions, and the way in which they combine. A selfish person need not be someone who lacks any disposition to help others. They may simply have a far stronger disposition to help themselves, so that the disposition to help others is rarely manifested. Thus another way in which it is possible to act out of character is to act from a disposition which is easily prevented from being manifested. Matching Principles allows for creditworthiness in such cases. For a nice example, consider Markovits' case of the fanatical dog-lover who 'performs a dangerous rescue

²² The counter-examples thus have more purchase on the view that creditworthiness, or virtuous action, requires acting from a settled disposition. And indeed, this is their more usual target (Hurka 2006; see Cokelet 2015 for discussion).

operation to save a group of strangers, at great personal risk' (2010: 210). Markovits argues that the dog-lover is creditworthy even though he wouldn't have saved the strangers if his dog had needed help at the same time. Matching Principles agrees: the dog-lover's act manifests a disposition to respond to the needs of others, even though this disposition could easily be blocked. For the same reason, Matching Principles allows the stingy person and the cowardly person, in the above examples, to be creditworthy, so long as they are acting from appropriate dispositions.²³

3) A third objection is that the view is not demanding enough:

Suppose that George is disposed to follow the principle, 'if you promise to ϕ , then ϕ '. On the assumption that it's a normative principle that, in general, the fact that you promised to ϕ is a reason to ϕ , it follows from Matching Principles that George is creditworthy for keeping his promises. But he might not be. Suppose, for instance, that he is only disposed to keep his promises because he is also disposed to do things which he thinks will get him into heaven, and he thinks that keeping his promises will get him into heaven. In that case, George is not morally creditworthy. (He might be prudentially creditworthy).²⁴

Cases like this show that we need to take more care in specifying the dispositions which make for creditworthiness (as already noted in n.19). There are different options here. One option is to note that George is disposed to keep his promises only

²³ See Mantel forthcoming b for further helpful discussion of these issues.

²⁴ Kant's shopkeeper example can be understood in a way that parallels this case. We can understand the shopkeeper as acting from the principle 'if x is the right change, give the customer x ', which is plausibly a correct moral principle. But he only acts from this principle because he thinks that doing so is good for business. See Mantel forthcoming b for this point, and relevant discussion.

as a consequence of a broader disposition: roughly, a disposition to do things he thinks will get him into heaven. We might thus distinguish between *basic* personal-level dispositions and *consequential* personal-level dispositions, and claim that it is the former which matter for creditworthiness. Another option is to deny that George is following the principle ‘if you promised to ϕ , then ϕ ’. Rather, George is following the principle, ‘if you promised to ϕ , and doing what you promised to do will help you get into heaven, then ϕ ’. Since this principle does not match an explanation of why the fact that George promised to do something is a reason to do it, following this principle does not make for creditworthiness.

These responses raise another concern. It might be thought that a basic disposition to ϕ when you have promised to ϕ is just a final desire to keep your promises. If so, then the view seems to differ little from the sophisticated version of the Matching Reasons view defended by Julia Markovits (2010). On Markovits’ view, creditworthiness requires that the *final* reasons for which you act match *final* reasons to act. For instance, if you have a final desire to keep your promises and you ϕ because ϕ -ing will keep a promise, this fact is your non-instrumental reason. You are then creditworthy if keeping promises really is worth doing for its own sake (and ϕ -ing does keep a promise). This looks much like what we have described as acting from the principle of keeping one’s promises.

I accept that the view of creditworthy action defended here is close to Markovits’ view. However, in several respects the view defended here is more general and so, I think, more satisfying and illuminating. First, we need not assume that a basic disposition to ϕ when you have promised to ϕ is a final desire to ϕ . For instance, some opponents of the Humean theory of motivation reject this assumption. One way the view defended here is more general than Markovits’ view is thus that it

does not assume the Humean theory. Second, and more significantly, the view applies to reasons for belief and emotions too. Markovits' view does not, insofar as the final/instrumental distinction she appeals to applies only to reasons for action, and action-related attitudes (such as desire and intention). Thus even if the view of creditworthy action defended here is equivalent to Markovits' view, it subsumes it under a more general picture.

4) Here is a more general worry that the view makes creditworthiness too easy to come by:

The problem with Matching Reasons was that it can be accidental that your motivating reasons match normative reasons. But can't it also be accidental that your motivating principles match normative principles? And if it is, isn't it then accidental that you do the right thing? And how can you then be creditworthy?

I think it's not entirely clear what this objection comes to. As noted, creditworthiness seems incompatible with it being too much of an accident that you responded to a normative reason. But as we have also seen, if you are motivated by normative reasons and normative principles, you could not have quite easily have failed to respond to a reason. Normative principles ensure a kind of reliability: at least other things equal, normative principles, correctly applied, will lead you to respond to normative reasons.

Nor is it clear what it would be for our motivating principles to only accidentally match normative principles. The idea can't be that our principles

accidentally match normative principles when they're not held for the right reasons, since motivating principles aren't the sort of thing which are held for reasons.

The thought might be that your motivating principles accidentally match correct principles when you could quite easily have followed principles which didn't match normative principles. However, it is not clear that the fact that you might easily have followed other principles precludes creditworthiness. For example, suppose that you were brought up so as to keep your promises and be honest but that you could easily have been adopted by people who would bring you up to lie, cheat, and steal. It surely doesn't follow that you are not creditworthy when you manifest your dispositions to keep your promises and be honest.

Still, one might reasonably think that our principles have to meet some kind of explanatory conditions. For instance, one might worry that principles which are brought about through indoctrination or implantation do not make for creditworthiness, even if they match normative principles. Thus one might suggest that our principles must be brought about, or sustained, or integrated, in the right sort of way.²⁵ This idea requires more discussion than I can offer here, and I don't want to rule it out; those who are sympathetic to it will take it to show that Matching Principles remains incomplete.

5) A final objection is that the view rests on controversial assumptions:

The view assumes that ethics and motivation are *principled*: that particular reasons hold in virtue of general principles, and that particular responses to reasons manifest more general dispositions. But these assumptions are

²⁵ For relevant discussion, see Arpaly 2002: ch.4; Bratman 2000; Fischer and Ravizza 1998: ch.7.

disputed. For instance, particularists in ethics deny that particular reasons hold in virtue of general principles. We should reject a view of creditworthiness with these commitments.

There is some truth in this objection but I think it's overstated. Although not entirely ecumenical, the view can be accepted by a wide-range of theorists. For example, the view does not assume universal normative principles – i.e. principles of the form: whenever it's the case that p , that p is a reason to ϕ . It is compatible with 'hedged' principles – principles of the form: generally/defeasibly/other things equal, when it's the case that p , that p is a reason to ϕ (cf. McKeever and Ridge 2006: ch.6, Horty 2012: ch.6, Väyrynen 2006). That's because the principles we follow might be similarly hedged. Thus the view is compatible with some forms of particularism in ethics.

It is true that the view rules out more extreme forms of particularism – forms which hold that particular reasons do not hold in virtue of even hedged principles. But I'm not sure that we should expect a view of creditworthiness to be entirely neutral on substantive moral questions. Furthermore, it does seem unclear how extreme forms of particularism can make sense of creditworthiness. We have seen that Matching Reasons, which is compatible with these versions of particularism, is unsatisfactory. And we have not seen how to improve on this view without assuming that there are at least hedged normative principles.²⁶

7. Conclusion

²⁶ For discussion of the related question of how particularists can make sense of competent moral judgment, see Dancy 1993: 63-64, 2004: ch.8.

I have argued that it is not enough for creditworthiness that your motivating reasons match normative reasons. Your motivating principles must also match normative principles. This view remains an outline. Much more needs to be said, for example, about what it is to act from a principle, what it is for one's principles to approximate to normative principles, and the relationship between pro tanto and overall creditworthiness. Nonetheless, I hope what I have said here is enough to make these further issues seem worth pursuing.²⁷

References

- Alvarez, M. 2010. *Kinds of Reasons*. Oxford: Oxford University Press.
- Arpaly, N. 2002. *Unprincipled Virtue*. Oxford: Oxford University Press.
- 2006. *Merit, Meaning, and Human Bondage*. Princeton, NJ: Princeton University Press.
- Arpaly, N and Schroeder, T. 2014. *In Praise of Desire*. Oxford: Oxford University Press.
- Bratman, M. 2000. Fischer and Ravizza on Moral Responsibility and History. *Philosophy and Phenomenological Research* 61 (2): 453-8.
- Broome, J. 2013. *Rationality Through Reasoning*. Oxford: Blackwell.
- Carroll, L. 1895. What The Tortoise Said To Achilles. *Mind* 4 (14): 278-80.
- Cokelet, B. 2015. Dispositions, Character, and the Value of Acts. In C. Miller et al (eds), *Character: New Directions from Philosophy, Psychology, and Theology*. Oxford: Oxford University Press, 233-50.

²⁷ Thanks to Alex Gregory, Chris Howard, Susanne Mantel, Conor McHugh, Kurt Sylvan, Daniel Whiting, and audiences in Berlin, London, Nottingham, Southampton, and Tucson. Work on this paper was supported by the Arts and Humanities Research Council [grant number AH/K008188/1].

- Crisp, R. 2000. Particularising Particularism. In B. Hooker and M. Little (eds), *Moral Particularism*. Oxford: Oxford University Press, 23-47.
- Dancy, J. 1993. *Moral Reasons*. Oxford: Blackwell.
- 2000. *Practical Reality*. Oxford: Oxford University Press.
- 2004. *Ethics Without Principles*. Oxford: Oxford University Press.
- Fischer, J.M. and Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Gregory, A. Forthcoming. Might Desires Be Beliefs About Normative Reasons? In J. Deonna and F. Lauria (eds), *The Nature of Desire*. Oxford: Oxford University Press.
- Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.
- Hills, A. 2009, Moral Testimony and Moral Epistemology. *Ethics* 120: 94-127.
- Horty, J. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.
- Hurka, T. 2006. Virtuous Act, Virtuous Dispositions. *Analysis* 66 (289): 69-76.
- Kant, I. 1785 [1997]. *Groundwork for a Metaphysics of Morals*, ed. M. Gregor
Cambridge: Cambridge University Press.
- Kearns, S. and Star, D. 2009: Reasons as Evidence. *Oxford Studies in Metaethics* 4:
215-242.
- Lord, E. Forthcoming. On The Intellectual Conditions for Responsibility: Acting for
the Right Reasons, Conceptualisation, and Credit. *Philosophy and
Phenomenological Research*.
- Lord, E. and Sylvan, K. Manuscript. Prime-Time (For the Basing Relation).
- Mantel, S. Forthcoming a. Worldly Reasons: An Ontological Inquiry into Motivating
Considerations and Normative Reasons. *Pacific Philosophical Quarterly*.
- Forthcoming b. Three Cheers For Dispositions: A Dispositional Approach to
Acting for a Normative Reason. *Erkenntnis*.

- Markovits, J. 2010: Acting for the Right Reasons. *Philosophical Review* 119: 201-242.
- 2012. Saints, Heroes, Sages, and Villains. *Philosophical Studies* 158 (2): 289-311.
- McGrath, S. 2011. Skepticism about Moral Expertise as a Puzzle for Moral Realism. *Journal of Philosophy* 108 (3): 111-137.
- McKeever, S. and Ridge, M. 2006. *Principled Ethics*. Oxford: Oxford University Press.
- Millar, A. 1991. *Reasons and Experience*. Oxford: Oxford University Press.
- Railton, P. 2005. Normative Guidance. *Oxford Studies in Metaethics* 1: 3-34.
- Raz, J. 2000. *Engaging Reason*. Oxford: Oxford University Press.
- Scanlon, T.M. 1998. *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.
- Schroeder, M. 2015. Knowledge is Belief for Sufficient (Objective and Subjective) Reason. *Oxford Studies in Epistemology* 5: 226-52.
- Sliwa, P. 2016. Moral Knowledge and Moral Worth. *Philosophy and Phenomenological Research* 93 (2): 393-418.
- Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
- Star, D. 2008. Moral Knowledge, Epistemic Externalism, and Intuitionism. *Ratio* 21 (3): 329-43.
- 2015. *Knowing Better*. Oxford: Oxford University Press.
- Sturgeon, N. 1988. Moral Explanations. In G. Sayre-McCord (ed), *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, pp. 229-55.
- Stratton-Lake, P. 2000. *Kant, Duty, and Moral Worth*. London: Routledge.
- Turri, J. 2010. On the Relationship Between Propositional and Doxastic Justification. *Philosophy and Phenomenological Research* 80: 312-326.

Väyrynen, P. 2006. Moral Generalism: Enjoy in Moderation. *Ethics* 166: 707-41.

Wedgwood, R. 2006. The Normative Force of Reasoning. *Nous* 40 (4): 660-86.