

Moral Worth & Knowing How to Respond to Reasons

J.J. Cunningham

University of Nottingham

Jesus College, University of Oxford

There is a difference between doing the right thing and being creditable for doing so. It's right to avoid buying goods from retailers who rely on sweatshop labour. But if one does so merely because one wants to fit-in with one's ethically conscious friends, one cannot be morally credited for doing the right thing. It's right to keep a promise to a friend. But if one does so merely because one fears gaining a bad reputation if one reneges, one cannot be morally credited for doing the right thing either. What grounds the difference between merely acting rightly and being morally creditable for doing so?

Part of the answer is that the agent must do the right thing out of a morally laudable motive. But even this is not sufficient for morally creditable action. That's because one could do the right thing for a morally laudable motive even though it's merely *an accident* that one does the right thing, given that motive. What this brings out is that there is a *non-accidentality constraint* on morally creditable action: acting in a morally creditable way requires a non-accidental fit between the agent's morally laudable motive and the rightness of their act. But what does fulfilling this non-accidentality condition consist in? There is a live dispute amongst moral philosophers about that question. This paper is a contribution to that dispute.

There are two main views of morally creditable action currently available. One is the *Right Making Features View*, according to which morally creditable action consists in action motivated by the very features of the act which make it right.¹ The other is the *Rightness Itself View*, according to which morally creditable action consists in action motivated by the fact that the act is right.² I aim to show that neither of these views successfully captures the non-accidentality condition. The problem with the first

¹Arpaly (2002, 2004) and Markovits (2010, 2014) endorse different versions of this view.

²The Rightness Itself view is defended by Herman (1996), Johnson King (2020), and Sliwa (2015, 2016).

is that one non-accidentally does the right thing only if one manifests a minimally correct understanding of the normative significance of the features of the action which move one to do it. The main problem with the second is familiar: one can non-accidentally do the right thing even if one is disposed to deny that the act is right. Neither of these problems is new. But here I will give them fresh development and defence.

In light of these failures I will defend and develop an alternative: the *Know How View*. According to the Know How View morally creditable action is action which is motivated by features of the act which make it right, where one's being motivated in this way is mediated by the exercise of one's *knowing how* to respond to those feature.³ On this view, not just any old motivation by right-making features will do: one's action needs to manifest one's knowledge of the significance of those features. But on this view, morally creditable action needn't involve motivation by rightness itself, for one can know how to respond to right-making features, and manifest that knowledge in action, even if one is disposed to deny that one's action is right. After sketching the Know How View and explaining how it captures the non-accidentality condition better than its opponents I develop the view in an attempt to undermine the objection that it collapses into a problematic form of Reliabilism about moral credit. According to the development I'll recommend, knowing how to respond to a reason is an ability which can be manifest in action *only if* the act is a *correct* response to the reason in question.

In §1 I isolate the target of dispute more precisely: morally creditworthy action and the non-accidentality condition thereon. In §2 I'll examine the two main views of morally creditable action and find them wanting. In §3 I explain the Know How View and show that it does better than the previously criticised alternatives. In §4 I reply to the Reliabilism objection to the Know How View.

1 The Target: Morally Creditable Action

Let's begin by pin-pointing our target phenomenon: morally creditable action – action for which the agent deserves credit for doing the right thing. Consider first:

The Egoist. Jasmine goes to a friend's dinner party. When the point in the evening is reached at which the other guests are suitably inebriated, Jasmine sneaks off to the kitchen and steals £10 from the coat of one of the guests, motivated by the desire to buy a bottle of wine on the way home.

³Obviously, this view bears a similarity to the view of Lord (2017, 2018). See footnote 25 for how my view differs from Lord's, however. It might also be thought to bear striking similarity to the view of Mantel (2018). See §§4.1–4.3 for an explanation of how my view differs from Mantel's.

Clearly, Jasmine's action is not morally permissible and so we cannot say that Jasmine deserves credit for doing the right thing. Morally creditable action requires that the action has a certain positive moral status: it is *right*.⁴ Now consider:

The Lucky Egoist. Dylan is also a guest at the dinner party. As the guests are preparing to leave Dylan sees that one of them is upset because £10 has gone missing from their coat. He goes to comfort them and suggests that the guests do a whip-round to help recoup the lost £10. Little do the other guests know, however, that Dylan is motivated by a desire to impress his upset friend, who he has had a crush on for some time – he doesn't particularly care that they are upset.

Although Dylan does the right thing in trying to help the victim of Jasmine's theft, Dylan doesn't act in a morally creditable way. That's because Dylan doesn't act out of a morally laudable motive: like Kant's shopkeeper (1996 [1785]), he does the right thing but he's motivated by self-interest. Morally creditworthy action thus requires both that the *action* has a positive status (rightness) and that the *agent* does too: their motive reflects well on them from the moral point of view.

But is it enough for morally creditable action that the agent does the right thing out of a morally laudable motive? It isn't. Consider a third case:

The Lucky Altruist. Leo is due to attend a charity gala this evening. He's going because he (justifiably) believes that the charity will help raise money in the fight against homelessness – a morally laudable motive. In fact, unbeknownst to Leo, the people running the event have no intention of using the money for that purpose. Instead, they intend to simply embezzle it. Fortunately for Leo there's something else he doesn't know: he's forgotten that he promised his colleague he'd meet them at the gala. He goes, successfully discharging the promise, but not in fact successfully promoting the goal of fighting homelessness.

Leo does the right thing: he promised to attend the charity gala, and that makes it right for him to go. He also acts with a morally laudable motive: it's a perfectly beneficent thing, to be motivated by the needs of the homeless. But he doesn't act in a morally creditable way because it's *an accident* that Leo's motive leads him to do the *right* thing. Leo goes because he believes that going will help in the fight against homelessness. But

⁴For the purposes of this paper 'right' can be read as a stand-in for any thin concept of positive moral appraisal.

because that consideration is false, it isn't what makes it right. Rather, what makes it right is that he promised to go. And yet Leo does not go for *that* reason.

Morally creditable action thus involves (at least) three ingredients: (i) the agent's act has the positive status of *rightness*; (ii) the agent has the positive status of being moved to act by a *morally praiseworthy motive*; and (iii) there is a non-accidental fit between these two elements. This third condition is the *non-accidentality condition*, which any account of moral credit is going to have to explain.⁵

One further point is in order. It's plausible that morally creditable action is an instance of a more general normative phenomenon: *normative achievement*. Normative achievement is the general phenomenon of an agent being creditable for doing as they ought. Normative achievement comes in as many flavours as there are flavours of ought: an agent can be creditable for doing the *prudent* thing, for doing the *epistemically required* thing, for doing what is desirable from the *aesthetic* point of view, for doing what is *rationally* required... and so on. This general phenomenon exemplifies the tripartite structure advertised. Morally creditable action is plausibly just the moral instance of this general phenomenon. Any account of moral credit should aim to respect that point by being generalisable to normative achievements across the board.⁶

2 Two Views of Morally Creditable Action Rejected

In this section I'll examine the two main views of morally creditable action and argue that each fails to adequately explain the non-accidentality condition. §2.1 criticises the Right-Making Feature View (henceforth: RMF); §2.2 criticises the Rightness-Itself View (henceforth: RI). This clears the ground for the development of my own position in §§3–4.

The refutation of each of these views and the development of an alternative which is supposed to avoid the problems for each is nothing new in the literature on moral worth.⁷ But I hope to offer fresh developments of extant problems for RMF and RI, and will be noting along the way where those developments appear.⁸

⁵In framing the problem of morally creditable action in terms of accounting for the non-accidentality condition I follow Sliwa (2016), Johnson King (2020), Howard (forthcoming), and Singh (Forthcoming).

⁶A point acknowledged by certain of my opponents, including: Sliwa (2016: §8) and Johnson King (2020: §4).

⁷See, for example, Mantel (2018), Lord (2017), Way (forthcoming), Howard (forthcoming), and Singh (Forthcoming).

⁸In particular, see footnotes 14 and 19.

2.1 The Right-Making Features View

When an action is right, it has certain features which make it right. Blowing the whistle on one's corrupt boss is right *because* it stops one's boss from exploiting his workers and shareholders and rescuing the cat from the burning building is right *because* she will die a painful death otherwise. According to the Right-Making Features View, morally creditable action consists in action which is motivated by the very features of the action because of which it's right.

Arpaly (2002, 2004) and Markovits (2010, 2014) present the view in terms of *normative reasons*. The normative reasons for one to ϕ are the facts of one's situation which speak for (or against) one's ϕ -ing. When the act is right, its rightness is explained by certain features of the action. These features are the normative reasons to perform the action which make it right. RMF can thus be understood as saying that morally creditable action consists in action motivated by the normative reasons which make the action right.

What RMF amounts to exactly will depend on what account of motivation by normative reasons one opts for. I'm going to build-into RMF Arpaly and Markovits's understanding of motivation by normative reasons. According to this account, action motivated by a normative reason only requires action motivated by belief in the relevant reasons⁹ plus some appropriate motivational state in virtue of which the agent ascribes non-instrumental value to acts of the type in question. Acting motivated by a normative reason thus does not require the agent to register the normative significance of their reasons. It does not, for example, require the agent to believe that the relevant features are normative reasons.

Building this view of responding to normative reasons into RMF yields the view that morally creditable action itself doesn't require the agent to represent their action as right, and nor does it require the agent to represent their reasons as right-making features, as moral reasons, or even as normative reasons of any kind at all.¹⁰ According to RMF: All that's required of the agent's motivational structure as far as moral credit is concerned is that it counts as motivation by reasons which are *in fact* the right making ones, where this need only involve a series of beliefs in the facts that are one's moral reasons plus a morally laudable non-instrumental concern for actions with the feature in question.

Given all this, RMF can most perspicuously be formulated in the following way:

Right-Making Features View. (i) Morally creditable action consists in ϕ -ing motiv-

⁹Perhaps subject to certain epistemic constraints: see, for example, Hornsby (2008); Hyman (1999, 2015); and [REDACTED].

¹⁰Arpaly (2002: 226), Markovits (2010: 208).

ated by the normative reasons for one to ϕ which make ϕ -ing right. But where:
(ii) ϕ -ing motivated by a normative reason only requires belief in the reason and a morally laudable non-instrumental concern for performing actions with the feature associated with that reason. (iii) So that: ϕ -ing motivated by a normative reason does not require the agent to register the normative significance of that reason.¹¹

RMF has several advantages. First, it gives us a natural cashing out of the non-accidentally condition: “[a]ctions motivated by right-making reasons... are not merely accidentally right. If I am motivated by right-making reasons, it is no coincidence that my motive issues in the right action” (Markovits, 2010: 211). Second, RMF gives us the right verdicts concerning the trio of cases we looked at above. About The Egoist, RMF correctly implies that Jasmine isn’t subject to moral credit: her action isn’t right in the first place, after all, so there are no right-making reasons for her to act in the light of. About The Lucky Egoist, RMF correctly implies that Dylan’s action isn’t subject to moral credit either: his reasons for action are not those which make it *right*, but rather those which make it *prudent*. And about The Unlucky Altruist, RMF correctly implies that Leo isn’t subject to moral credit: the considerations from which he acts, although morally laudable ones, are not identical to the normative reasons which make actually going to the gala right. And finally, RMF readily generalises to normative achievements across the board: the proponent of RMF can simply say that, in general, for an agent to be creditable for responding as they ought is just for them to respond in a way that is motivated by the normative reasons which make it the case that they ought to respond in that way. Moral credit is just the moral instance of that phenomenon.¹²

Nevertheless, RMF is subject to counterexample. Consider:

Bad Moral Reasoner. Lucy remembers that she promised her partner she’d do the school run this morning. But she also has an early morning meeting which, though not particularly important, it’d be mildly awkward to miss, and which clashes with the school run. She doesn’t deliberate about what to do, because she’s been in this situation several times before and feels she knows immediately what to do: she’ll do the school run because she promised to. But Lucy is in fact seriously confused about the ethics of promise-keeping. Were she to explicitly reason about what to do, she’d

¹¹Although Arpaly (2002) and Markovits (2010) each endorse this view, their views of moral credit differ in more specific ways. For example, Arpaly (2004) accepts, whereas Markovits (2010: 210) denies, that there is a counterfactual requirement either on moral credit itself or on degrees of moral credit.

¹²A further standard motivation for RMF is the case of Mark Twain’s (2017 [1884]) Huckleberry Finn, first discussed in the philosophical literature by Bennett (1974). I postpone discussion of this until §2.2.

think this:

I promised I would do the school run, and that's important. But it's generally correct to prioritise work meetings, even if it'd be only mildly awkward to miss them, over promises to family. However, I can ignore that requirement on me to prioritise work as long as discharging the promise would also make me happy, which it will – it means I get to spend the morning with my kids, which is rare these days; so: I'll do the school run.

Her decision to do the school-run manifests that implicit, and deeply incorrect, understanding of the significance of promise-keeping.

Lucy does the school-run and she does it motivated by the fact that she promised: her doing it manifests the belief that she promised plus a non-instrumental concern for doing what she promises. We can suppose, moreover, that her promise does indeed outweigh whatever professional obligation she has to attend the unimportant meeting, and so the fact that she promised really does make it right for her to do the school run. Thus, she does the school-run for the reason that makes it right and she meets the criteria laid down by RMF for morally creditable action.

But the way her concern for promise-keeping moves her to action is *via* triggering an ethically objectionable incompetence with the kind of moral reason for which she acts. Her action manifests a disposition to give improper weight in her reasoning to the fact that she promised. She treats this sort of reason as outweighed by mild professional concerns, and the overriding force of these as in turn undercut when the promise carries prudential value – and we can safely assume that the normative significance of promise keeping does *not* pattern in this way in the practical reasoning of agents who are even moderately skilled with considerations of promise keeping. She does end up acting for the reason that she promised. It's just that that consideration moves her *via* a grossly defective understanding of its normative significance. It thus seems too much like she merely *accidentally* does the right thing. RMF therefore fails to capture the non-accidentality condition.¹³

In response, the proponent of RMF could remind us that it's a requirement of their theory that the agent's action manifests not just a belief in the relevant right-making reasons, but also a pro-attitude *which is itself morally laudable*. This isn't an *ad hoc*

¹³Even if Arpaly's counterfactual requirement could help with this counterexample, that requirement is far too strong – as Markovits (2010: 210) and Sliwa (2016: 7) point out and as the case I consider in §4.2 demonstrates.

feature of their theory, moreover. Morally creditable action is defined as involving: (i) action which is right; (ii) which manifests a *morally praiseworthy* motive; and (iii) which is non-accidentally right given that motive. That second aspect of morally creditable action is itself morally loaded: what's required is not just motivation by any old motive, but by one which is itself morally laudable, and the proponent of RMF captures this by making it a requirement of responding to right-making reasons that the pro-attitude the agent manifests is a morally laudable one. The proponent of RMF could then suggest that Lucy's pro-attitude towards promise-keeping is not itself morally laudable because it embodies a deeply flawed understanding of the normative significance of that practice. So Lucy doesn't meet the criteria laid-down for morally creditable action by RMF after all.¹⁴

But this response to the problem won't do. The response assumes that we cannot separate our ethical assessment of Lucy's pro-attitude towards promise-keeping from our ethical assessment of her disposition, which her action manifests, that embodies her understanding of the ethical significance of promise-keeping. The latter is certainly morally incorrect. The response we're considering effectively has it that it follows from this that the former is morally incorrect too. But that's too strong: what seems right is that it's perfectly morally laudable for Lucy to non-instrumentally value keeping her promises, it's just that, on this occasion, her *way* of acting motivated by that value is morally faulty because mediated by her deeply incorrect understanding of how promise keeping weighs against competing considerations. The current reply does not speak to that concern.¹⁵

Might the proponent of RMF try instead insisting that Lucy *is* morally creditable for keeping her promise, but only to a *low degree*? The problem with this move is that it's plausible that Lucy displays *no* degree of moral credit for doing the right thing. After all, Lucy flagrantly misunderstands the normative significance of promise keeping and her action manifests this deep misunderstanding. Given this, she fails the non-accidentality condition, which is a condition on moral credit *simpliciter* and not moral credit *to some degree*.

In saying all this, I'm presupposing that some individuals can fail to understand the normative significance of some class of considerations to such a grave degree that a response to a consideration of that kind on their part might only accidentally be in-line with what the consideration favours. Of course, my opponent might not find it intuitive

¹⁴Compare the similar cases offered by Sliwa (2016: 6), Way (forthcoming: §2), and Singh (Forthcoming: §3), none of whom consider the response just broached.

¹⁵On a similar note, as Way (forthcoming: 6–7) points out, it won't help to suggest that Lucy's reason is more properly identified with the content of her hypothetical deliberation: that would require us to implausibly conflate Lucy's reasons with the principles of reasoning she relies on.

that *Lucy* falls into that category. But it's surely plausible that it *is* possible for some agent to fall into it, and that's all I need, it seems to me, to clinch the point against RMF.¹⁶

What's gone wrong? Because RMF doesn't require the agent to manifest any kind of sensitivity to the normative significance of their right-making reasons, it leaves it open that the agent might act motivated by those reasons but in a way which also manifests a deeply incorrect understanding of their normative significance. But as cases like Bad Moral Reasoner attest: such agents fail to satisfy the non-accidentality condition. The upshot would appear to be that RMF should be abandoned in favour of a theory of moral credit which requires the morally creditable agent to manifest a grip on the morality of their action. Let's now turn to such a theory.

2.2 The Rightness Itself View

The main competitor to RMF is the Rightness Itself View. According RI:

Rightness Itself View. Morally creditable action consists in ϕ -ing motivated by the fact that ϕ -ing is morally right, where this (at least) involves the agent correctly believing that the action is morally right and that being amongst the reasons that motivate them to ϕ .

There are different versions of RI available. According to Johnson King (2020): "An act has moral worth just in case it is an instance of someone *deliberately* doing the right thing" (*Ibid.*: 14). And according to Sliwa (2016): "A morally right action has moral worth if and only if it is motivated by concern for doing what's right (conative requirement) and by knowledge that it is the right thing to do (knowledge requirement)" (*Ibid.*: 2). I will focus largely on the generic version of the thesis, above, in what follows.

RI gives us another plausible-looking account of the non-accidentality condition: if an agent ϕ s in a morally creditworthy manner, then, according to RI, the very fact of its moral rightness is amongst the agent's reasons for doing it – so it's *no accident* that the action is in fact morally right, given the way the agent is motivated. Moreover, RI gives us the correct verdicts about the trio of cases looked at in §1. About The Egoist, RI correctly implies that Jasmine isn't subject to moral credit: her action isn't morally right in the first place, so that very fact cannot be her reason for doing it. About The Lucky Egoist, RI correctly implies that Dylan's action isn't subject to moral credit either: his reasons don't include the fact that the action is *morally right* – at best, they include only the fact that his action is *prudently right*. And about The Unlucky Altruist,

¹⁶Thanks to my reviewer for pressing me on this point.

RI correctly implies that Leo isn't subject to moral credit: the considerations from which he acts, although morally laudable ones, do not include the fact that the act is morally right.

Moreover, RI has an important advantage over RMF: it correctly predicts that Lucy, in *Bad Moral Reasoner*, doesn't act in a morally creditable manner. After all, Lucy doesn't do the school run on the basis of the belief that doing so is morally right. Rather, she goes in the light of her belief that she promised, without deliberation.¹⁷

Should we therefore go ahead and endorse it as our account of morally creditable action? I don't think so. There are two problems for it I want to highlight.

First, we should note that RI doesn't plausibly generalise to normative achievements across the board. A generalised RI would say:

Rightness Itself View Generalised. S being creditable for ϕ -ing as they (epistemically, prudentially...) ought consists in S being moved to ϕ by the fact that they (epistemically, prudentially...) ought to ϕ , where this involves (at least) the agent correctly believing that they (epistemically, prudentially...) ought to ϕ and that being amongst the reasons that motivate them to ϕ .

But this generalised version of RI is implausible when it comes to epistemic achievements. When agents form the beliefs that they epistemically ought to hold in response to the evidence, they don't in general do so moved partly by the thought that the weight of evidence makes a decisive case for the belief. Such beliefs are often formed immediately in the light of the evidence, with the rarefied thought that they ought to hold the belief in question coming, if at all, only in moments of self-critical reflection, which might never arise. Nevertheless, beliefs correctly held in the light of the evidence often constitute normative achievements. Similarly, the generalised version of RI yields implausible results when it comes to the emotions. When agents respond emotionally to their environment, their response is often immediate and not mediated by the thought that the response is the one they ought to opt for. At best, agents tend to retroactively normatively endorse their immediate emotional responses during self-critical reflection. And yet, it seems perfectly plausible that agents often respond with apt emotions and that they are often creditable in doing so.

This failure to generalise is not a *decisive* problem for RI: it's a disadvantage of the view which should be weighed against the disadvantages of competing pictures. But there is a more decisive problem for RI which I now want to discuss. Consider the extensively discussed case of Twain's (2017 [1884]) *Huckleberry Finn*:

¹⁷Isn't RI false because it morally credits agents who act out of rightness but on a false conception thereof (see for example Singh (Forthcoming: §4))? Sliwa's knowledge requirement handles such cases.

Huckleberry Finn. Huckleberry Finn is a teenager growing up in the American deep south during the middle of the 19th century. He's internalised the racist ideology of his sub-culture: he believes that slavery is morally permissible, that slaves count as being owned by their masters, and hence that helping to set a slave free constitutes stealing – something he acknowledges to be seriously morally objectionable. Huck has, however, befriended the fugitive slave Jim. He feels sympathy towards Jim, in spite of his slave status, and when the opportunity arises, he chooses to lie to the authorities to help Jim escape, something about which Huck feels profoundly conflicted.

Finn's act of helping Jim escape is the morally right thing to do. But is Finn creditable for doing the right thing? And is Finn motivated by the moral rightness of the act itself?

Arpaly (2002) and Markovits (2010) offer the following description of Finn's motivational structure: Finn judges that he ought not help Jim because such a thing is morally precluded by the fact that helping Jim constitutes stealing. But he is motivated to help Jim by the fact that Jim is a person plus his laudable non-instrumental concern for Jim's personhood. The fact that Jim is a person is what makes it right for him to help, so that Finn does the right thing and does it for the reason which makes it right, even though he's disposed to describe what he does as immoral.

Relative to this description, Arpaly and Markovits argue that Jim isn't motivated by the fact that his act is right – indeed, he's precisely disposed to deny that. But they also think it intuitive that Finn is creditable for doing the right thing. That's because it'd be churlish of us to deny moral credit to agents like Huck who find themselves, through historical misfortune, in a culture which has imbued them with an incorrect sense of right and wrong, but who nevertheless manage to attune themselves emotionally and motivationally to the actual demands of morality. Such agents act in morally creditable ways in spite of the incorrect moral code they inhabit. To the extent that such a phenomenon is intuitively possible – and I think it is – we will find it plausible that Huck acts in a morally creditable manner, so it's suggested.

If these verdicts are right, we have a counterexample to RI. However, I don't think the proponent of RI should be convinced, as things stand. The problem is that Finn's action is no more morally creditable than Lucy's in Bad Moral Reasoner. Lucy's action is not morally creditable because her action manifests a deeply incorrect understanding of the ethical significance of promise-keeping. But Finn's act of helping Jim also manifests a deeply incorrect understanding of the morality of personhood: he treats the fact that

Jim is a person as outweighed by the consideration that Jim is the property of someone else. It's as if, were Finn to explicitly deliberate, he would reason like this: "Jim's a friend but he's also a slave – he *belongs* to Miss Watson. I really shouldn't be doing this at all. But I can't help myself – I'll do it anyway". And this reasoning clearly manifests a deep mistake about the significance of Jim's personhood: that fact *does* defeat the consideration that Jim belongs to Miss Watson. So if Lucy's action lacks moral worth, and I've argued that it does, then by the same token Finn's action lacks moral worth. The Arpaly-Markovits reconstruction of Finn's moral psychology does not give us a counterexample to RI after all.

All is not lost, however. Consider the following twist on the case:

Reflective Huck. During the process of helping Jim escape, Huck takes time to reflect on his actions, and he comes to a more sophisticated understanding of himself and his ethical relation to his social milieu. He decides that his act of helping Jim to escape *is* wrong and that *from the moral point of view*, he shouldn't be helping. But he recognises Jim to have the status of a person, and that this consideration is a decisive reason to help Jim – it's just not a *moral* reason. Huck concludes that morality isn't always the most important thing – sometimes, one should break the moral rules, and this is such an occasion.

In this twist on the standard scenario, Finn's experiences lead him to a relatively sophisticated understanding of his action. In Finn's view it's an act which is justified, but not morally so. In fact, his act *is* morally right and he acts motivated by the reasons which make it so, it's just that he isn't prepared to describe those considerations as *moral* ones and so he isn't prepared to call his action *morally right*. But he does correctly treat those considerations as defeating whatever countervailing considerations there are to turn Jim in. Clearly, in this variant, Finn still isn't motivated to help Jim by the fact that it's morally right to do so: he's still disposed to explicitly deny that. *But* Finn does treat the fact of Jim's personhood correctly: he treats it as defeating the consideration that he belongs to Miss Watson. His only mistake is that he's disposed to describe that consideration as non-moral. But the fact that he's not prepared to describe these considerations as moral doesn't mean that he's failed to understand their significance – his correct understanding of their significance is manifested in his reasoning. It's therefore plausible, I submit, that Finn's act *is* morally creditworthy, even though not motivated by the fact that the act is morally right.¹⁸

¹⁸Oughts come in many different flavours: moral oughts, prudential oughts, epistemic oughts... and so

How might the proponent of RI respond to this counterexample? Johnson King (2020: §2) argues that:

Johnson King’s Principle: “For all properties of acts *F*, someone accidentally does an *F* thing [in the sense pertinent to morally creditable action] if she has no idea that her act possesses property *F* when she performs it” (*Ibid.*: 3).

It follows from this principle that an agent counts as non-accidentally doing the *morally right* thing only if they believe that the act is morally right. If Johnson King’s Principle is independently plausible, my counterexample fails.

But Johnson King’s Principle is itself subject to counterexample:

Renegade Painter. Johan is highly skilled at portraiture. His portraits are regarded by experts as being of exquisite beauty. But Johan is a peculiar artist. He thinks that only paintings of nature can truly be described as *beautiful*. He thinks that while what he does is of some value, he denies that it’s of value from the point of view of beauty and he’d deny that he has any skill in producing beautiful paintings.

Suppose Johan paints a beautiful portrait and his painting the portrait manifests his high level of skill in doing so. Then we’d surely want to say that he’s creditable for producing a beautiful portrait and hence that his action is *non-accidentally* one of producing a beautiful portrait. But Johan doesn’t believe that what he produces is beautiful, such is his misunderstanding about the property of beauty. Thus, the sense of ‘accident’ pertinent to an agent being creditable for performing an act with property *F* cannot be one on which an agent accidentally does an *F*-thing if they have no idea that their act has property *F*. We should therefore reject Johnson King’s Principle.¹⁹

on. Is there also such a thing as ought *simpliciter*, which the interplay of these specific kinds of lower-level oughts determine? *Contra* Baker (2018), I think there *is*. Moreover, that commitment is part of Reflective Huck: Reflective Huck precisely thinks that although he *morally* ought not help Jim, he ought to do so *simpliciter*. The proponent of RI takes no official stance on the existence of ought *simpliciter*. But they have the option of responding to my case by dropping their neutrality and actively denying the existence of ought *simpliciter*. I do not have space to defend the existence of oughts *simpliciter* here. But I do think the coherence and intuitive plausibility of cases like Reflective Huck, as well as cases like Williams’s (1976) Gauguin, serve to shift the onus of proof onto those who deny its existence. I thank my reviewer for pressing me on this.

¹⁹Further counterexamples to Johnson King’s Principle are furnished by Lord’s (2017: §§4.1–2) parables of Sammy and Marcel and Howard’s (forthcoming: §4) football and music examples. Neither Lord nor Howard show sensitivity to the fact that the standard description of the case of Huck fails to undermine RI, however, so that a variant such as Reflective Huck is required.

3 The Know How View

Where does all this leave us? What we need is an account of morally creditable action which enables us to correctly diagnose Bad Moral Reasoners as failing to act in a morally creditable manner but which does so in a way that's compatible with the idea that agents like Reflective Huck, who are disposed to deny that their reasons for action are moral ones, nevertheless do act in a morally creditable manner. It needs to do all this, moreover, whilst preserving our intuitions about the three basic cases presented in §1: The Egoist, The Lucky Egoist, and The Unlucky Altruist, and whilst being readily generalisable to normative achievements across the board. This section presents my own view of morally creditable action (§3.1), which I argue meets these *desiderata* (§3.2): the *Know How View*.

3.1 The Basic Idea

According to the Know How View (henceforth: KHV):

Know How View. (i) Morally creditable action consists in ϕ -ing motivated by the normative reasons which make ϕ -ing right. But where: (ii) ϕ -ing motivated by a normative reason, p , of a certain type, T , consists in ϕ -ing in a way that manifests one's knowing how to respond to reasons of type T , where that ability is triggered by one's knowledge of the fact that p . And: (iii) In the case of morally creditable action, the right-making reason is of a type because of which it is right-making, and the agent ϕ s in a way that manifests their knowing how to respond to reasons of *that* type, triggered by their knowledge of the particular right-making reason at issue.

Let me explain. We can distinguish between different types of normative reason: reasons of beneficence, reasons of courage, reasons of honesty, reasons of prudence, aesthetic reasons, epistemic reasons... and so on. Agents can be more or less skilled in their dealings with normative reasons of a certain type. Someone who is disposed to ascribe far more importance to their own well-being than to the well-being of others is someone who lacks competence with respect to both reasons of prudence and reasons of beneficence. Someone who is disposed to treat considerations of promise-keeping as always trumped by considerations of beneficence similarly lacks competence with reasons of promise keeping. Someone who is disposed to treat their own freedom as having no importance is incompetent with respect to reasons of autonomy... and so on.²⁰

²⁰For the general idea that agents can be more or less skilled in their responses to reasons see also Annas

As long as one possess a certain degree of skill with respect to one’s dealings with reasons of the relevant type, one might count as *knowing how* to respond to reasons of that type. Knowing how to respond to a reason of a certain type doesn’t require that one is a virtuoso with respect to reasons of that type: one can know how to respond to promise-keeping considerations even if there are situations in which one is not in a position to tell what promise-keeping requires. But there is nevertheless a minimal degree of competence one will have to exemplify to count as having the relevant piece of know-how, which the agents in the examples given in the preceding paragraph don’t meet. Knowing how to respond to reasons of a certain type involves being disposed to treat certain considerations as reasons of that type, for a range of responses (actions, intentions, emotions, doxastic responses...), and to do so *knowledgeably*.

KHV proposes to identify morally creditworthy action with action which is motivated by the normative reasons which make the action right. That is what clause (i) says. But KHV adds to this that we are to understand acting in response to a normative reason in terms of know how: to act in response to a reason of a certain type just is to act in a way that manifests one’s knowing how to respond to reasons of that type, where that ability is triggered by the agent’s propositional knowledge of the particular normative reason at issue.²¹ That’s what clause (ii) says: it elaborates clause (i) by providing a know-how based theory of what it is to act motivated by a normative reason.

But how exactly does this idea apply in the case of morally creditable action? Clause (iii) of KHV answers that question: in the case of morally creditable action, the right-making reason one correctly responds to will be of a type – a reason of promise keeping, a reason of courage, a reason of beneficence... – where *because* it belongs to that type it counts as right-making, in the context at issue.²² According to KHV, morally creditable action is action which manifests one’s knowing how to respond to reasons of *that* type. Where this piece of know-how is triggered on the occasion at issue by the agent’s

(2011), Mantel (2018), Lord (2017, 2018), and Howard (forthcoming). I should also flag a methodological point: I do not try to provide anything like a systematic account of what it is to know how to respond to reason of a certain type here. Instead, like the other philosophers just mentioned, I simply proceed by relying on our intuitive verdicts about cases.

²¹Notice here I build into KHV the idea that ϕ -ing correctly in response to the normative reason that p requires knowing that p. For defence of that epistemic constraint see Hornsby (2008) and Hyman (1999, 2015). For criticism see Hughes (2014) and Locke (2015). For a response to those critics see [REDACTED].

²²The caveat ‘in the context at issue’ is important. In saying that the type to which a normative reason belongs explains why it is right-making I do not mean to imply that the type to which a normative reason belongs *suffices for* or provides a *complete explanation of* its right-making status. A complete explanation of why the reason is right-making will include: the absence of relevant undercutting-defeaters; a complete specification of the counter-veiling reasons and their weight; a complete specification of the relevant intensifiers and attenuators; the presence of any motivational or epistemic constraints on reason-hood... and so on. My thought is only that the type to which a reason belongs can explain *relative to those contextually determined background conditions* why the reason makes the act is right.

propositional knowledge of the fact which is their reason.

Take an illustrative example. Mary hears a cry for help from next door, and thereby knows that her next door neighbour is in need of help. Mary is a mature ethical agent and knows what to do when someone is in need: she rushes round to see what the matter is, which is the right thing to do. Steve, who lives across the street, also hears the cry for help and thereby knows that the person is in need. Steve also goes to check on them, but only because he wants to be seen to be doing the right thing. Mary is creditworthy for doing the right thing; Steve isn't. KHV explains why: Mary's action manifests her knowing how to respond to reasons of beneficence, where the right-making reason she thereby counts as correctly responding to is right-making precisely because it is a reason of this kind. Steve's action doesn't manifest that ability.

KHV provides the same style of account of morally creditable action as RMF. Both views identify morally creditable action with acting motivated by right-making normative reasons. But crucially, KHV differs from RMF in the account it gives of what it is to correctly respond to normative reasons. For RMF, an agent may count as acting correctly motivated by a normative reason even if they in no way register the normative significance of their reason. But KHV requires that agents count as correctly acting for reasons only if they manifest their knowing how to respond to a reason of the type in question. And so KHV implies that there is a sense in which responding to reasons requires the agent to register the normative significance of their reason. But KHV *doesn't* require the morally creditable agent to believe that their act is right, like RI does – *that* kind of sensitivity to the normative status of one's reasons isn't required.²³

3.2 The Non-Accidentality Condition Explained

KHV provides us with yet another straightforward cashing-out of the non-accidentality condition on morally creditable action. According to KHV, the agent who is creditable for doing the right thing does the right thing *non-accidentally* because they do it in a way that manifests their skill in responding to reasons of the right-making type. The skill in question is possessed of a degree high enough for it to count as the agent knowing-how to respond to reasons of that type. It's no more an accident that the morally creditable agent does the right thing than it's an accident that the skilled pianist succeeds in playing the fugue the sheets for which they are following; the skilled hairdresser succeeds in cutting their customer's hair into the latest style; or our painter Johan succeeds in creating a

²³A fully developed account of morally creditable action will also account for *degrees* of moral credit. KHV naturally accommodates this as follows: the degree to which an agent is morally creditable is a function of the degree of know-how they manifest in responding to the reason of the type at issue. I aim to develop this aspect of KHV in future work.

beautiful portrait (even though he's disposed to deny that that's what he's done).

We considered a trio of basic cases in §1. KHV handles these perfectly well. About The Egoist, KHV correctly predicts that Jasmine isn't morally creditable: her action isn't right, so she can hardly count as being motivated by right-making reasons. About The Lucky Egoist, KHV also gives us the correct verdict: Dylan does the right thing, but at best his action counts as manifesting his knowing how to respond to reasons of prudence. And finally KHV gives us the correct verdict about The Lucky Altruist: Leo doesn't count as manifesting his knowing how to respond to the reasons of the type which make the action right because he doesn't count as manifesting his knowing how to respond to reasons of promise-keeping.

Moreover, KHV provides us with an account of moral credit which is readily generalisable to other normative achievements:

Know How View Generalised. (i) Responding as one ought, in a way for which one is creditable, consists in responding motivated by the normative reasons which make it the case that one ought to respond in that way. But where: (ii) Responding motivated by a normative reason, p , of a certain type, T , consists in responding in a way that manifests one's knowing how to respond to reasons of type T , where that ability is triggered by one's knowledge of the fact that p . And: (iii) in the case of being creditable for doing as one ought, the ought-making reason is of a type because of which it is ought-making and the agent responds in a way that manifests their knowing how to respond to reasons of that type, triggered by their knowledge of the ought-making reason at issue.

To illustrate, let's think about what this generalised version of KHV would amount to when it comes to the normative achievement of responding emotionally appropriately to a work of art. In this sort of case, we have three elements: the features of the artwork which constitute aesthetic reasons to respond emotionally in the way one does; the aesthetically correct response in question – the response favoured by those reasons, from the aesthetic point of view; and a suitable link between these two elements so that the agent's correct response is a success attributable to her. The generalised version of KHV gives us the following story: the emotion is an aesthetic achievement on the agent's part because it manifests the agent's knowing how to respond to aesthetic reasons, where that ability is one which is triggered by the agent's knowledge of the salient features of the artwork and it's because the artwork has those features that the response is correct, from the aesthetic point of view.

So far so good. But what about the cases which proved a problem for RMF and RI? Let's start with Bad Moral Reasoner. Lucy remembers that she promised her partner

she'd do the school-run, even though it clashes with a not-so-important work meeting, and goes ahead and does it for the reason that she promised. Her doing so manifests gross incompetence with reasons of promise-keeping. And for that reason, her act isn't morally creditable. KHV straightforwardly gives the correct verdict about Lucy. Her misunderstanding of the normative significance of promise-keeping is indeed so bad, we should deny that she *knows how* to respond to such considerations.

What about Reflective Huck? Here's how KHV handles the case. Reflective Huck knows how to respond to reasons of personhood: he knows that such reasons outweigh reasons of stealing; by-and-large he's disposed to feel the right range of emotional responses to Jim given Jim's status as a person; he's disposed to feel guilt if he failed to help Jim, because Jim's a person; and he's disposed to recommend to others in a similar social position to him to do the same thing were they to find themselves in his circumstances, again because of Jim's personhood. Huck's understanding of the morality of personhood is no doubt defective in a number of ways, but he qualifies as knowing how to respond to reasons of personhood nevertheless: knowing how to do something, as already remarked, doesn't require one to be a virtuoso with respect to it – one can know how to ride a bike even if in particularly difficult terrain one would lose control of it. But given that Reflective Huck knows how to respond to reasons of personhood, why not say that his act of helping Jim manifests that ability? And if we do say that, of course, KHV would give us the correct verdict: it would imply that Reflective Huck is creditable for doing the right thing.

The main grounds for doubting that KHV secures us this result about Reflective Huck is this: Since Huck believes that the reason of personhood he has to help Jim is not a moral reason, Huck doesn't really *know how* to respond to reasons of personhood. In other words: Huck's incorrect belief concerning the moral status of his reason implies that Huck is disposed to treat his reason of personhood in ways that would preclude him displaying the degree of competence required for Huck to count as knowing how to respond to that type of reason in the first place. If this is right, KHV doesn't secure us the intuitively correct result that Huck is morally creditable for helping Jim.²⁴

My response to this objection comes in two parts. First, I think we should note that the onus of proof is on the proponent of the objection. After all, it *seems* perfectly coherent to describe the case of Reflective Huck as one in which Huck knows how to respond to reasons of personhood and as manifesting that ability in his response to Jim, even though he's resistant to thinking of his reason as a moral one. Moreover, cases like Renegade Painter, presented at the end of §2.2, confer independent plausibility on the idea that it's possible for an agent to know how to ϕ under a certain description (for

²⁴My thanks to my reviewer for pressing me on this.

example, paint a beautiful artwork), and exercise that ability in ϕ -ing, even though they are incorrectly disposed to deny that their ϕ -ing falls under that description. But if we have independent reason to think that this is possible, we should expect an analogous thing to hold of reasons: it's possible for an agent to know how to use a particular kind of moral reason, and exercise that ability in using a reason of that kind, even if they are incorrectly disposed to deny that the reason is a moral one.

These considerations show that the onus of proof is on the proponent of the present objection to show that Huck's disposition to deny that his reason is of moral significance precludes him knowing how to respond to reasons of personhood. How might the proponent of the objection argue for that claim? The most plausible way for them to do so, it seems to me, is to argue that because Huck is disposed to deny that his reason for helping Jim is a moral reason, it must be that he doesn't fully understand what morality is. But then, so the current argument goes, he also doesn't fully know how to respond to moral reasons, including reasons of personhood: there are contexts in which Huck's partial ignorance of what morality is means that he's disposed to incorrectly respond to a reason of personhood.

But in order for this argument to achieve the desired result, it would have to be that one knows how to ϕ only if one *fully* knows how to ϕ . That is: one knows how to ϕ only if one is not disposed to make any mistakes in how one ϕ s. But that's surely too strong: knowing how to ϕ doesn't require one to be a virtuoso, let alone infallible, with respect to ϕ -ing. Examples abound: one can know how to play the piano even if one is not exceptional at it; one can know how to make a good break-off shot in snooker even if one doesn't always succeed in doing it; one can know to replace a tyre on a car even if there are some esoteric models of car with respect to which one wouldn't have a clue... and so on. I conclude, then, that the challenge to the current objection to my KHV-friendly account of what's going on in Reflective Huck stands.

4 The Spectre of Reliabilism

I've outlined KHV and explained how it provides us with a satisfactory account of the non-accidentality condition. I now want to consider an objection to KHV: that it collapses into an unsatisfactory kind of Reliabilism. I will argue that a response to this objection requires us to accept the *Success Constraint*: knowing how to F can manifest itself only if its manifestation is a successful instance of F . In §4.1 I articulate the Reliabilist view at issue. In §4.2 I explain why the view is problematic and I explain how this generates a problem for KHV. In §4.3 I motivate the Success Constraint and explain how it solves the problem.

4.1 Reasons-Reliabilism Introduced

Consider the following view:

Reasons-Reliabilist View. (i) Morally creditable action consists in ϕ -ing which is motivated by the normative reasons which make ϕ -ing right. But where: (ii) ϕ -ing motivated by a normative reason, p , of a certain type, T , consists in ϕ -ing in a way that manifests a disposition to treat p as a T -reason to ϕ , which serves to reliably produce responses that are correct responses to T -reasons, and where that disposition is triggered by one's knowledge of the fact that p . And: (iii) In the case of morally creditable action, the right-making reasons are of a type because of which they are right-making, and the agent ϕ s in a way that manifests a reliable disposition to respond to reasons of that type.

The Reasons-Reliabilist View (henceforth: RRV) is structurally parallel to KHV. The only difference is that it cashes out motivation by normative reasons in a way that's at least notionally distinct from KHV: it identifies motivation by normative reasons with action that manifests a certain kind of disposition possessed by the agent. This disposition is a disposition to treat certain considerations as normative reasons of a certain kind for certain responses. Moreover, the disposition is *reliable*: the disposition serves to render the agent's various responses to perceived reasons of the relevant type liable to be a correct response to those reasons.²⁵

In my view, the most well-developed version of RRV is to be found in Mantel (2018). Consider, for example, the following passage, which supplies us with the rudiments of Mantel's account of successfully responding to a normative reason:

Agent A performs action X for the normative reason that p if and only if... Agent A performs action X; There is an entity, that p , and it is a normative reason for A to do X; A has the belief that p ... [and] [t]he action X is a manifestation (with respect to the normative reason that p) of the agent's competence to do what is favored by the normative reasons of a certain fairly (including the normative reason that p)... [Where] [t]his competence consists in three sub-competences, namely of the *epistemic competence* to represent the normative reasons... the *volitional competence* to be motivated by these descriptive beliefs to do what is favored by the represented reason, and the *executorial competence* to execute these motivations. (*Ibid.*: 43)

²⁵In footnote 28 I suggest that Lord's (2017; 2018) view is ambiguous between RRV and KHV.

According to Mantel, we're to analyse correctly acting for a normative reason in terms of the manifestation of a collection of reliable dispositions: the 'epistemic competence', which reliably functions to ensure that the agent's belief in their reason is reliably correct; the 'volitional competence', which reliably functions to ensure that the agent is motivated to act in-line with what their type of reason favours; and an 'executorial competence' which reliably functions to ensure that the agent performs an action, at the right time, which matches the content of their motivation. Correctly responding to a normative reason consists in acting in a way that manifests these three dispositions in-tandem. Moreover, Mantel (2018: Ch.4) puts this account to work in providing an account of moral worth: for Mantel, acting in a morally worthy way consists in correctly responding to a right-making normative reason, where this is understood in terms of the account above. I think it's clear from all this that Mantel endorses (a version of) RRV.

26

That's the basic idea of RRV and I've identified at least one proponent of it in the literature. However, in order to fully cash-out the view, we'd have to have some explanation of what reliability amounts to. There are various options here, and, to my mind, Mantel herself simply isn't clear on which notion of reliability she wants to go for. I will now fill-in this lacuna in RRV's position with the suggestion that I think is most natural.

In my view, the proponent of RRV's best bet would be to opt for a modal notion of reliability, modelled on the safety condition from the epistemic setting. According to this idea, intuitively: a disposition to respond to reasons of a certain type is reliable just in case not easily could that disposition lead S to respond incorrectly to reasons of the type at issue. More precisely:

Reasons-Safety. A disposition, *D*, to treat certain considerations as normative reasons of type *T* for various responses is reliable if, and only if: Were the agent to respond in a way that manifests *D*, S's response would count as a correct response to the consideration S's knowledge of which triggers *D*.

In the possible worlds idiom: such a disposition is reliable just in case at the nearest worlds at which S responds by manifesting that disposition, at those worlds S also counts as responding correctly to the consideration which moves them. Putting this together with RRV we get the view that morally creditable action consists in action motivated by right-making reasons, but where the latter consists in manifesting a disposition such that not easily could the agent be moved to respond for the same type of reason and respond incorrectly.

²⁶For criticism of Mantel's picture, see [REDACTED].

4.2 Reasons Reliabilism & the Non-Accidentality Condition

RRV fails to handle the non-accidentality condition. Consider the following:

Masochist. Daisy has a housemate, John. Daisy hears a cry of pain coming from John's room, and thereby knows that John is in pain. She treats the known fact that John is in pain as a normative reason of beneficence to go to help, and thereby does so. Little does Daisy know, however, that John is actually a masochist, who enjoys pain when it's voluntarily inflicted on him, and that is indeed what is happening in the present situation. Daisy is therefore incorrect, on this occasion, to treat the known fact that John is in pain as a reason to go to help.

The fact of someone else's pain is a reason to help them only if the person in question does not enjoy being in pain. Since that condition isn't fulfilled in Masochist, Daisy doesn't really have a normative reason to help John.

But now suppose we develop the case as follows:

Masochist II. Daisy has a second housemate, Tom, who's room is next door to John's. Tom knows about John's masochistic tendencies, knows that the latter's cry of pain isn't cause for moral concern, but knows that Daisy is on her way up the stairs motivated by beneficence. Tom sneaks into John's bedroom, out of sight of John, and fires a pellet gun at him, thereby making John cry in pain, but this time involuntarily. Tom does this because he knows that Daisy's coming to help out of beneficence, and he thinks it'd be funny to make her decision to help correct after all.

In this case, the fact that John is in pain goes from not being a reason for Daisy to help to being a reason for Daisy to help, given Tom's actions. Throughout Daisy's process of action, her act manifests her disposition to treat the known fact that John is in pain as a reason of beneficence to help. Prior to Tom's interference, Daisy was incorrect to treat that fact as having that normative significance, and so trivially doesn't count as correctly responding to the *normative reason* that John is in pain. Post Tom's interference, however, she's *correct* to treat the known fact that John is in pain as a normative reason to go help. Post Tom's interference, does Daisy count as successfully responding to the normative reason that John is in pain?

Intuitively, no: again, it appears to be merely a matter of *luck* that Daisy's action accords with the moral reason at issue. It's merely a matter of luck, as far as Daisy's motivation is concerned, that her action ends up being one for which her motive is a

normative reason to do what she does. To really bring this point home, note that Daisy doesn't have any reason to go to help *prior to* her being motivated by beneficence to do so. Thanks to Tom's actions, she only actually has a reason to go help *because* she's already decided to help on the basis of beneficence. But that's surely the wrong way round, if we're to count Daisy as genuinely responding to the normative reason to help, after Tom's interference.

What is the proponent of the Reasons-Reliabilist View committed to saying about Masochist II? I think they're committed to saying that, after Tom's interference, Daisy *does* act motivated by the right-making reason that John is in pain, and is therefore worthy of moral credit. That's because Daisy's act of coming to help John manifests a disposition which is reasons-safe: it's a disposition to treat a range of considerations – *that they are in pain, that they are being robbed, that they are being lied to* – as reasons of beneficence for various responses. And we can take it that at nearby worlds at which Daisy acts in a way that manifests that disposition, she acts correctly: she's a well-functioning moral agent, who just so happens to find herself in an odd situation.

It'll be immediately objected that there *is* a nearby world at which Daisy treats the fact of John's pain as a reason of beneficence to help, but it is no such thing: a world at which Tom hears what's going on, but doesn't interfere. But we can easily build details into the case which enable us to see Tom's act as itself counterfactually robust: we can build it into the case that Tom is a home-bird who's always in his room, who loves his practical jokes, and who plays jokes of that nature all the time ... so that were Daisy to treat the known fact that John is in pain as a reason of beneficence to help, Tom would also play his practical joke.

The lesson is that it doesn't suffice for successfully responding to a normative reason that one acts in a way that manifests a modally robust disposition to act correctly in response to reasons of a certain type. That might be so even if, on the occasion in question, it's merely an accident of one's external circumstances that one acts correctly and that the disposition one manifests continues to count as reliable.

That's the counterexample to the Reasons-Reliabilist View. The problem for the Know How View can now be easily stated: isn't there only merely a verbal difference between the Know How View and the Reasons-Reliabilist View? What's the difference between saying that moral credit consists in manifesting one's *knowing how* to respond to reasons of the right-making type and saying that it consists in manifesting a *disposition to reliably respond* to reasons of the relevant type? And if there is no difference between these, KHV runs into the above counterexample.

4.3 The Success Constraint

To combat this problem, I suggest we combine KHV with:

The Success Constraint. Necessarily, if S ϕ s in a way that manifests their knowing how to F then S's ϕ -ing counts as a successful instance of type F .

Let me clarify what this thesis amounts to before arguing that it's plausible. Pieces of knowing how are constituted by dispositions possessed by the relevant agent. As Schellenberg (2018: Ch.2) point out, we can distinguish between the *success condition* associated with a disposition and its *manifestation conditions*. The manifestation conditions of a disposition are just the characteristic events or properties which are brought about upon the disposition being triggered. The success condition of a disposition obtains just in case the disposition manifests *well* or as it *should*. For many dispositions, its manifestation conditions are not identical to its success condition: it can manifest *unsuccessfully*. For example, one might think the disposition to visually experience red surfaces in appropriate lighting conditions can manifest even if one doesn't *see* the red surface. Given this background, we can identify the Success Constraint with the view that when a disposition constitutes a piece of *know how*, its manifestation conditions don't come apart from its success condition: it can manifest only ever successfully.²⁷

With that background in place, let me give some illustrative examples in order to elucidate the Success Constraint further. According to the Constraint: If the pattern of movements displayed by your fingers on the piano really counts as manifesting your *knowing how* to play the fugue then you succeed in playing the fugue; if your conversation manifests your *knowing how* to persuade people then your conversation is a successful instance of persuading them; if the way you respond emotionally to a victim's account of what happened to them manifests your *knowing how* to empathise then you're emotional response is a correct piece of empathy... and so on. The Success Constraint generalises. It says that *whenever* something manifests one's knowing how to F : that thing is an instance of F as it was aimed to be insofar as the agent exercised their knowing-how to F – it's an instance of F which satisfies whatever success condition we associate with the disposition that constitutes the piece of know-how in the first place.

To deny the Success Constraint would be to say that one can ϕ in a way that manifests one's knowing how to F but where one's ϕ -ing doesn't count as a successful instance of F . On this view, one could move one's fingers over the piano keys in a way that manifests

²⁷Explaining the Success Constraint this way assuages the worry that the Constraint is trivial, because all dispositions to F are such that they can manifest only if F obtains. This is trivial only if the disposition is being individuated in terms of its manifestation conditions – that is, only if 'F' denotes the manifestation conditions of the disposition, and not its success condition.

one's knowing how to play the fugue even if one doesn't count as successfully playing the fugue: one makes enough mistakes to warrant the judgement that the fugue doesn't count as having been successfully played at all. And on this view, one could respond emotionally to the victim in a way that manifests one's knowing how to empathise with them even if one's emotional response is inappropriate.

The Success Constraint is plausible. First, I think it gives us the intuitively correct verdict about the cases I've briefly described: there is something very odd about saying the pianist *manifests her knowing how to play the fugue* without her successfully playing it – the correct thing to say would be that she *attempts* to display her knowledge but fails. *Mutatis mutandis* for the attempt at empathy. So the Success Constraint should have the status of the default view about knowing how.

Moreover, the Success Constraint gives us a plausible answer to the question of why we should count knowing how to *F* as an instance of *knowing*. Presumably, we do so only because there is something in common between knowing how and knowing that. But what? The Success Constraint provides us with an answer: knowing how and knowing that are both mental states which are essentially successful. Knowing that *p* is essentially successful because it entails the presence of a belief which is as it ought to be: it's a true belief, for example. Knowing how to *F* is essentially successful because it can manifest itself only if its manifestation is a successful instance of *F*. The Success Constraint is a thesis about knowing how which is a natural counterpart of the thesis of the factivity of knowing that.

So the Success Constraint is an independently plausible view about knowing how. If we combine the Success Constraint with KHV, we get the following picture. KHV says that morally creditable action consists in action which is motivated by the normative reasons which make it right. But we're to think of acting motivated by a normative reason as acting in a way that manifests one's knowing how to respond to reasons of the type under which the normative reason in question falls, where that ability is triggered by the known normative reason at issue. When the agent acts in a morally creditable way, their normative reason falls under a type because of which that reason is right-making, and their action manifests their knowing how to respond to normative reasons of the type at issue. Combining this with the Success Constraint yields the view that the know-how which is manifested by the agent's action is one which could be manifested at all *only if* the agent's action is a successful response to reasons of the right-making type at issue: *only if* the action is a correct response to the particular consideration of promise-keeping, beneficence, personhood...or whatever. That's because the success condition of knowing how to respond to a reason of a certain type consists in a manifestation which matches – really is favoured by – the reason to which one responds. This means

that KHV gives a particularly robust cashing out of the non-accidentality condition: it implies that morally creditable action is action which is non-accidentally right because it is action which manifests an ability such that the action is *guaranteed* to be a successful response to the right-making reason, if it counts as manifesting that ability at all.

Moreover, if we combine KHV with the Success Constraint, this serves to put distance between KHV and RRV. For the proponent of RRV, it's metaphysically possible for the agent to manifest their reliable disposition to respond to reasons of the type at issue even if their response is incorrect. Because the disposition is *reliable*, at the *nearest* worlds at which the agent responds in a way that manifests that disposition, the response will count as correct. But nevertheless, there will be *some* non-near worlds at which the agent responds in a way that manifests their disposition and their response is *incorrect*. According to KHV, supplemented by the Success Constraint, such a thing is metaphysically impossible: since the disposition the agent who correctly responds to normative reasons manifests is *essentially successful*, there is *no* possible world, near or far, at which the agent responds by manifesting *that* disposition and their response is incorrect.

Finally, if we combine KHV with the Success Constraint, KHV avoids the Maoschist II counterexample which beset RRV. With the Success Constraint on the scene, the proponent of KHV can simply deny that Daisy manifests her knowing how to respond to reasons of beneficence: Tom's act precludes that ability from manifesting itself (although it doesn't preclude it being possessed). It might *seem* to Daisy just like she's manifested her knowing how to respond to reasons of beneficence but, because of Tom's actions, she's mistaken. It emerges, then, that KHV should be thought of as quite different from a Reasons Reliabilist View and is in a much stronger position to explain the non-accidentality condition than the latter. Since it's also better in that respect than the Right-Making Reasons View and the Rightness Itself View, I hope to have made a strong case that it's the best theory of morally creditable action currently on the market.²⁸

²⁸Obviously, the view I have developed bears a similarity to the view of responding to reasons independently developed by Lord (2017, 2018). But there are significant differences. First, although Lord cashes out his view in terms of know how, it's equivocal whether he accepts what I've called the Success Constraint. He seems to clearly endorse it in Lord (2017). But in Lord (2018: §5.4.2) he argues that it *suffices* for manifesting knowing how to respond to a reason in ϕ -ing that one satisfies certain counterfactuals that look much like the counterfactuals the proponent of RRV analyses responding to normative reasons partly in terms of. Second, even if Lord is committed to the Success Constraint, he simply builds it into his position. That raises the question of *why* he should do so – why he shouldn't just opt for an account like RRV instead. I have tried to motivate the Success Constraint: both as an independently plausible constraint on knowing how in general and as something the proponent of KHV in particular needs to endorse in order to avoid collapsing into RRV. And finally, Lord motivates his know how based view of responding to reasons partly by appeal to the thought that it's the best way to avoid a variety of conceptualisation constraints which supposedly serve to make responding to reasons over-demanding. My KHV, and its attendant Success Constraint, is not motivated in that way – I've

References

- Annas, Julia (2011). *Intelligent Virtue*. Oxford: Oxford University Press.
- Arpaly, Nomy (2002). “Moral Worth”. *Journal of Philosophy*, 99 (5): 223. doi: 10.2307/3655647.
- (2004). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.
- Baker, Derek Clayton (2018). “Skepticism About Ought Simpliciter”. *Oxford Studies in Metaethics*, 13.
- Bennett, Jonathan (1974). “The Conscience of Huckleberry Finn”. *Philosophy*, 49 (188): 123–134. doi: 10.1017/S0031819100048014.
- Herman, Barbara (1996). *The Practice of Moral Judgement*. Camb. Mass.: Harvard University Press.
- Hornsby, Jenninfer (2008). “A Disjunctive Conception of Acting for a Reason”. In: *Disjunctivism: Perception, Action, Knowledge*. Ed. by Adrian Haddock and Fiona Macpherson. Oxford: Oxford University Press: 262–279.
- Howard, Nathan Robert (forthcoming). “One Desire Too Many”. *Philosophy and Phenomenological Research*. doi: 10.1111/phpr.12642.
- Hughes, Nick (2014). “Is Knowledge the Ability to ϕ for the Reason that p ?” *Episteme*, 11 (4): 457–462.
- Hyman, John (1999). “How Knowledge Works”. *Philosophical Quarterly*, 50 (197): 433–451.
- (2015). *Action, Knowledge, and Will*. Oxford: Oxford University Press.
- Johnson King, Zoë A. (2020). “Accidentally Doing the Right Thing”. *Philosophy and Phenomenological Research*, 100 (1): 186–206. doi: 10.1111/phpr.12535.
- Kant, Immanuel (1996 [1785]). “Groundwork of the Metaphysics of Morals”. In: *Practical Philosophy*. Ed. and trans. by Mary J. Gregor. Cambridge: Cambridge University Press.
- Locke, Dustin (2015). “Knowledge, Explanation, and Motivating Reasons”. *American Philosophical Quarterly*, (52) (3): 215–232.
- Lord, Errol (2017). “On The Intellectual Conditions for Responsibility: Acting for the Right Reasons, Conceptualization, and Credit”. *Philosophy and Phenomenological Research*, 95 (2): 436–464. doi: 10.1111/phpr.12306.
- (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.

therefore shown that the view can be motivated independently of such concerns. That’s a good thing, I hope to show in future work, because there is in fact good reason to think that the Know How View is committed to some variety of conceptualisation constraint.

- Mantel, Susanne (2018). *Determined by Reasons: A Competence Account of Acting for a Normative Reason*. New York, USA: Routledge.
- Markovits, Julia (2010). “Acting for the Right Reasons”. *Philosophical Review*, 119 (2): 201–242. doi: 10.1215/00318108-2009-037.
- (2014). *Moral Reason*. Oxford University Press.
- Schellenberg, Susanna (2018). *The Unity of Perception. Content, Consciousness, Evidence*. Oxford: Oxford University Press.
- Singh, Keshav (Forthcoming). “Moral Worth, Credit, and Non-accidentality”. In: *Oxford Studies in Normative Ethics*. Vol. 10.
- Sliwa, Paulina (2015). “Praise Without Perfection: A Dilemma for Right-Making Reasons”. *American Philosophical Quarterly*, 52 (2).
- (2016). “Moral Worth and Moral Knowledge”. *Philosophy and Phenomenological Research*, 93 (2): 393–418. doi: 10.1111/phpr.12195.
- Twain, Mark (2017 [1884]). *The Adventures of Huckleberry Finn*. William Collins.
- Way, Jonathan (forthcoming). “Creditworthiness and Matching Principles”. In: *Oxford Studies in Normative Ethics, Vol 7*. Ed. by Mark Timmons. Oxford University Press.
- Williams, Bernard (1976). “Moral Luck”. *Proceedings of the Aristotelian Society Supp.* Vol. 50 (226): 115–135. Reprinted in Williams (1981): 20-39. All page references are to the reprint.