Pre-Print. Forthcoming in Synthese.

Who's to (Instrumentally) Blame? Influenceability vs. Reasons-Responsiveness

Abstract

Blame is typically justified on the basis of retrospective desert. However, an emerging strand of account gives an alternative justification for blame: the forward-looking, or proleptic, effects of that blame in cultivating a desirable form of agency, shared moral considerations responsive agency. These instrumentalist accounts differ as to their grounding conditions: the agential features that licence blame in cases of moral failure. Some accounts advocate grounding such justified blame in terms of whether or not the agent meets the condition of influenceability; others advocate the condition of reasons-responsiveness. I will argue that influenceability is unsuccessful as a grounding condition: such accounts appear to licence blame in too many cases of bad action (or omission). In order to not do so, it is unclear how they can avoid either reducing to another kind of grounding condition, such as reasons-responsiveness, or abandoning any attempt to posit a substantive grounding condition. However, the reasons-responsiveness condition has also been attacked: it appears to be vulnerable to a mismatch between the proleptic goals of the accounts to which it is a part, and to empirical evidence suggesting an apparent situational lack of control. I will defend reasons-responsiveness as a grounding condition; however, I will suggest that instrumentalist accounts should be focused less on their grounding conditions and on blame, and more on the empirical efficacy of our reactive attitudes in scaffolding moral agency. In light of these considerations, I will suggest a friendly revision to Vargas' grounding condition drawing on Fischer and Ravizza, Holroyd, and Calhoun that, I think, better accommodates the proleptic concerns that motivate Vargas' and McGeer's accounts.

Keywords: instrumentalism, moral responsibility, control, blame, reasons-responsiveness, situationism

1 Introduction

An emerging strand of accounts of moral responsibility justifies the application of our prototypical reactive response, emotive blame, on the basis of its proleptic, or forward-looking, effects in cultivating a desirable form of agency, shared moral considerations responsive agency (SMCRA) (Fricker 2016; Jefferson 2019; McGeer 2019; Vargas 2013a). These instrumentalist accounts are specified in two ways. First, they are accounts of the *role* of blame (Vargas 2021). That is, they are not accounts of what unifies all instances of blame (as in, e.g., self-blame, blame

of the dead, blame of the absent), but instead attempt to capture the role of emotive blame in cultivating our SMCRA. Second, they *justify* the application of that blame on the basis of those proleptic effects.

One major point in dispute among these instrumentalist accounts is what *grounds* responsibility attributions. That is, while these instrumentalist accounts agree that, in general, blame is justified given that it serves those proleptic ends, they disagree about who the appropriate targets of moralized blame are. One camp takes the most straightforward approach (Jefferson 2019; McGeer 2015a, 2019): they claim that the appropriate targets of blame are those agents who are appropriately influenceable by that blame. For example, McGeer's Scaffolding View assesses an agent as being blameworthy just in case that agent has performed a bad action, and they maintain the temporally-extended skill-based capacity for sensitizability to moral considerations. Blaming agents who meet this condition, she claims, will serve to cultivate their SMCRA.

Her account, thus, is centrally distinct in two ways from the grounding conditions of the alternative camp, exemplified by Vargas' (2013a, 2013b, 2015, 2017, 2021) Agency-Cultivation Model (ACM). First, the accounts differ in their characterization of who the appropriate targets of blame are. By Vargas' (2013a: 200-204) lights blame is justified just in case an agent performs a bad action, and they are reasons-responsive, as well as maintaining any other necessary retrospective conditions on responsible agency (as in, e.g., the epistemic condition). For Vargas, maintaining reasons-responsiveness means, very roughly, manifesting the appropriate capacity to both detect and react to the salient moral considerations in that circumstance. Here, Vargas offers a permutation of standard reasons-responsiveness accounts of responsible agency (Fischer and Ravizza 1998; Nelkin 2011; Wolf 1990).

Second, they differ with regard to the level at which the justification for blame operates. On Vargas' account, token instances of blame are justified retrospectively, while the entire practice of moralized blaming is justified proleptically; by contrast, McGeer claims that token instances of blame are themselves individually justified proleptically. That is, by Vargas' lights, the cultivation of SMCRA operates at the level of the practice as a whole, whereas according to McGeer it operates over token instances of justified blame.

In this paper, I will contrast the grounding conditions of McGeer's Scaffolding View with those of Vargas' ACM. Both of their accounts are instrumentalist in these respects, and both are partially motivated by Situationist[1] evidence from social psychology. Why suppose one account is more appropriate than the other? Here, we can distinguish between two desiderata: the desert and proleptic desiderata.[2]

First, the desert desideratum: we want the account to give us a convincing retrospective grounding for attributions of responsibility. Note that while the desert desideratum is central to most standard compatibilist accounts

1 See §4.
2 In making this distinction, and in the solution that I offer, I am heavily indebted to Holroyd (2018).

of responsible agency, it does not have as great of a bite on either McGeer's or Vargas' instrumentalist accounts. This is because, while both accounts *do* attempt to accommodate the desert desideratum (McGeer 2019: 315-317; Vargas 2013a: ch.8), what ultimately matters for both accounts are the *effects* of blame in cultivating SMCRA.

So, second, the proleptic desideratum: given that each account ultimately justifies the application of blame on the basis of its capacity to appropriately cultivate SMCRA, we might take the account that better serves that cultivation to be superior. Note that standard compatibilist accounts do not need to appeal to the proleptic desideratum, and that making this desideratum centrally important renders instrumentalist accounts potentially vulnerable to empirical evidence regarding the actual efficacy of our reactive attitudes in cultivating SMCRA.[3]

On standard retrospective compatibilist justifications of the application of blame, all that needs to be appealed to are the theoretical and metaphysical considerations that undergird when agents meet the control condition, and any other conditions on responsible agency. By contrast, on a proleptically-justified account, it must *actually be the case* that the reactive attitudes we're applying serve the appropriate proleptic ends when that capacity is maintained. Otherwise, their application is, by the lights of the theory, unjustified. Now, it may not need to be the case that they are *always* effective in their proleptic ends; perhaps a preponderance of efficacy is sufficient. But the point is that, unlike standard retrospectively-justified accounts, proleptically-justified accounts need to take into account the empirical efficacy of our reactive attitudes in assessing the appropriateness of their grounding conditions for blame attributions.

In this paper, I will assess which type of grounding better serves the proleptic desideratum. I will ultimately conclude that Vargas' reasons-responsiveness condition better meets the proleptic desideratum than McGeer's influenceability condition. However, I will offer my own emendation of Vargas' ACM, which I argue will allow it to better meet the proleptic desideratum. I will offer two central conceptual points of departure from the grounding conditions of Vargas' ACM. First, I will claim that it is a mistake for Vargas to incorporate one form of epistemic consideration—awareness of the relevant moral considerations—into his grounding condition. This is because this consideration, while important given the desert desideratum, is, I claim, counterproductive in meeting the proleptic desideratum.

Second, I will claim that individuals are *apt targets* for the reactive attitudes when they maintain in that circumstance an appropriate responsiveness to what they take the relevant moral considerations to be. That is, on my suggested emendation, an assessment of responsibility is not a licence for blame, but merely suggests that an individual is responsible in that circumstance, and so some reactive attitudes and associated responses (which may not be blame) may be appropriate, depending upon the circumstances.

3 For such concerns as they apply to blame, see Eisenberg et al. 2006: 668-669; Kohn 2006; though see Brink and Nelkin 2022; see also Holroyd 2018; Moody and Nojoumian 2024; Pereboom 2021.

I make this move because I do not think blame is generally the best reactive response to deploy in order to serve our account's proleptic goals when individuals have performed a bad action and are apparently responsible. This is because I do not think that being blamed is a central explanation of our responsiveness to moral considerations—so focusing on blame diminishes an account's appropriateness in light of the proleptic desideratum (Holroyd 2018; Pereboom 2021).

In order to make this argument, I will first present, in §2, part of what motivates McGeer's adoption of influenceability as a grounding condition: the Hard Problem of Moral Responsibility (McGeer and Pettit 2015). The Hard Problem, McGeer suggests, is a problem for Vargas because it illustrates the worry that there may be a substantial conflict between the overall proleptic goals of Vargas' account with adoption of reasons-responsiveness as a grounding condition (McGeer 2015a).

I will then, in §3, sketch McGeer's alternative account given the Hard Problem. I will argue that McGeer's influenceability condition appears to license blame in *too many* cases of bad action. Licencing blame in too many cases of bad action will, I claim, likely not serve the proleptic desideratum. Avoiding this problem, I suggest, would mean either adopting something like Vargas' reasons-responsiveness grounding condition, or abandoning the attempt to proffer a substantive grounding condition.

However, as we saw above, the reasons-responsiveness grounding condition may be vulnerable to the Hard Problem. In §4, I will illustrate Vargas' response to the Hard Problem, and illustrate why the Hard Problem might be more of a substantive issue for reasons-responsiveness accounts than he supposes. I will then show how Vargas can respond to even this more substantive issue.

However, I will suggest that McGeer is right to worry that Vargas' account may ultimately create a mismatch between the set of subjects who are assessed, by Vargas' lights, as being blameworthy and the set that is appropriately blameworthy, given the proleptic desideratum—though not in the way she supposes, but instead due to the epistemic features baked into Vargas' reasons-responsiveness control condition.

Thus, in §5, I will sketch my suggested emendation of Vargas' ACM in order to respond to this issue. I will suggest a compromise account, inspired by Holroyd (2018), Calhoun (1989), and Fischer and Ravizza (1998), that seems to me to better fulfil the proleptic desideratum. And, for similar reasons, I will ultimately suggest that a focus on blame is, for instrumentalist accounts, misguided.

2 The Hard Problem of Moral Responsibility

McGeer and Pettit (2015) have identified what they call the Hard Problem of Moral Responsibility. The

Hard Problem is a problem for justifying the application of blame on the basis of an assessment of reasons-responsiveness, as Vargas suggests. According to the Hard Problem, purportedly blameworthy reasons-responsive agents never fully manifest their reasons-responsiveness capacities in the actual world; they're blameworthy, on the basis of such accounts, because it appears as though they would have been responsive to the relevant considerations in a sufficient number of nearby possible worlds. But why, McGeer and Pettit ask, should this show that they're blameworthy in *this* world? In *this* world, it might seem as though their lack of an appropriate responsiveness to presented moral reasons is, given their responsiveness in nearby possible worlds, the product of a brute factor: a fluke or a glitch.

So, for example, imagine that Henry neglects to help someone whom he's morally obligated to help, and whom he knows he ought to help. Reasons-responsiveness accounts would claim that Henry is blameworthy just in case in a suitable proportion of nearby possible worlds Henry would have been responsive to that moral consideration and so helped that person. But why, McGeer and Pettit ask, should this modal analysis of Henry's reasons-responsiveness capacity justify our *emotive condemnation* of him? Given Henry's assessed appropriate responsiveness in nearby possible worlds, his lack of responsiveness in *this* world would appear to be the product of a "glitch of some kind—say, a neural misfiring—[that] obstructed the operation of [Henry's] capacity, leading to the failure of response" (McGeer and Pettit 2015: 165). If Henry's action *is* the product of such a brute factor then not only does it seem questionable whether or not he's properly *blameworthy*, but it might even seem appropriate to *console* Henry, for his purportedly culpable action wasn't really his fault.

And the worry is that such a brute factor will always underlie blame attributions on standard compatibilist accounts of responsible agency; even if we attempt to explain Henry's culpability by reference to a trait he may possess such as laziness or akrasia, it remains the case that his being unable to resist or remove that trait will ultimately itself be down to a brute factor. As McGeer and Pettit (2015: 164) explain, "we might invoke a further non-brute factor to explain the failure to respond to the reasons *for* overcoming that trait. And while we might go on in the same vein, searching for ever deeper, non-brute explanations, it would make little sense to postulate an infinite regress of non-brute factors that account for the agent's failure in the case at hand. The explanatory regress must end somewhere by invoking a brute factor." The Hard Problem is that, given that Henry's action is, by analysis, ultimately the product of a brute factor, why should we emotively condemn him?[4]

4 In this respect, one might take the Hard Problem to also target Deep Self accounts of moral responsibility such as Doris' (2015) and Sher's (2009) for reasons similar to those posited by Wolf (1990): why hold agents to be responsible for action attributable to them if the reason that action is attributable to them is not ultimately down to them.

3 McGeer's Scaffolding View

The Hard Problem, thus, McGeer claims, should motivate us to abandon the attempt to utilize reasons-responsiveness as a grounding condition for instrumentalist accounts. McGeer's own account, the Scaffolding View,[5] developed over several papers, is itself geared in a manner, inspired by P.F. Strawson, to justify the specifically emotive character of blame in light of the Hard Problem.

As McGeer (2019: 316) explains:

> In thinking and (more likely) saying, "you could have done otherwise," we are rightly adopting a stance towards agents that honours their developmental potential. We are effectively telling them, "you have what it takes!"—"otherwise (suppressed premise) why would we bother taking you to task in this way?!" Of course, it needs to be added that *in so taking agents to task,* "you could have done otherwise!" is generally understood to have a perlocutionary force that goes beyond the mere reporting of some condition we take the relevant agents to be in (consider: "you could have done otherwise!" is rarely said in the calm cool tones of a philosopher reporting some abstract metaphysical possibility). Rather, we are *exhorting* agents to do better. And in so exhorting them, we thereby do our bit to sensitize agents to the reasons they earlier failed to track and thereby do our bit to help realize the developmental potential that our claim, "you could have done otherwise" attributes to them.

The thought here is that our own agency is itself largely scaffolded by the emotive blame directed at us. If we were not so regularly blamed, on her account, we would be substantially diminished in our SMCRA; it is that blame which *scaffolds* us.

There are, I think, two ways of understanding McGeer's instrumentalist account. According to the shallower reading, blame is justified given that it serves in general to scaffold that individual's future SMCRA. But if this is all that McGeer intends, there is no conflict between her account and Vargas'; she would be simply agreeing that blame is in general justified on the basis of its proleptic effects. But I don't think that this is all that McGeer intends. Instead, I believe that she intends a deeper account: to give an account of precisely what capacity would ground that justification. That is, she's not *just* saying that blame is justified given that it serves to scaffold moral agency, but giving an account of the *specific agential capacity* that, when it is met, renders blame appropriate.

But by McGeer's lights, given the Hard Problem, any appropriate account would not appeal to any atemporal modal capacity of the relevant individuals, as in, e.g., their reasons-responsive capacities. Instead, her account "calls for a fundamentally different kind of metaphysical analysis of such capacities: one that is explicitly dynamic, intertemporal and interpersonal rather than static, atemporal, and intrapersonal" (McGeer 2019: 312).

---

5 Note the contrast between McGeer's (2015a, 2019) proposed response to the Hard Problem, and McGeer and Pettit's (2015) positive response to the Hard Problem. By McGeer and Pettit's lights, blame may be proleptically justified on the basis of an assessment of reasons-responsiveness, given that that assessment implicitly invokes a moral audience. That is, in assessing your reasons-responsiveness and blaming you, McGeer and Pettit claim, I do not merely report the modal fact that in a nearby possible world you might have acted otherwise, but "I act in a consciously evocative manner, seeking to strengthen your standing sensitivity to the reasons I put before you or take to be before you" (McGeer and Pettit 2015: 181). By contrast, McGeer (2015a; 2019) claims that we ought to abandon any attempt to utilize an assessment of reasons-responsiveness as an appropriate grounding for blame attributions given the Hard Problem.

McGeer's Scaffolding View, she claims, motivates an account of the relevant agential capacities "essentially unlike the disposition-based model" (312) of standard reasons-responsiveness accounts.

Such accounts, McGeer claims, assume 'atemporalism'. She explains that "on this assumption, having a capacity (perhaps only relative to a particular range of considerations and circumstances) means having a *fully realized* capacity (again relative to that range of considerations and circumstance) *at the time* in question" (McGeer 2015a: 2646). But, by her analysis, given the Hard Problem, purportedly blameworthy agents will *never* have such a fully realized capacity.

So McGeer (2019: 315) proposes an alternative account of what feature of agents grounds their being appropriately held to be blameworthy.

> What they need to possess—indeed, all they need to possess—is a susceptibility to the scaffolding power of reactive attitudes, experienced as a form of moral address. Notice, first, that this is a substantially more modest requirement on responsible agency than was suggested by [reasons-responsiveness accounts]. For, on [those] views, at the time of their purported misdeed, agents must possess an atemporal, purely intrapersonal, disposition-based capacity for responding to the reasons. This, as we have seen, is a disposition to be in-the-moment responsive to the reasons present at the time of their action—a disposition that makes their failure to respond to those reasons both puzzling and apparently blameless. By contrast, the Scaffolding View holds that agents must possess an intertemporal and (essentially) interpersonal skill-based capacity for responding to the reasons—that is, a capacity that involves having whatever it takes to be *sensitizable* to the kind of reasons present at the time of their action, in part by way of the exhortatory effects of (ex post) reactive scaffolding.

By McGeer's lights, the only property of the agent which is assessed when we adjudicate whether or not to apply moralized blame is their temporally-extended skill-based capacity to be liable to the development of their receptivity to moral considerations via moralized blame, i.e., their influenceability, and an assessment of whether or not the moral considerations that they had (or had not, in the case of culpable omissions) responded to are desirable or not.

But this remains vague. Precisely which agents are blameworthy on her account? McGeer suggests in several places that we should want to *expand* the set of appropriately blameworthy agents. She claims, e.g., that for instrumentalist accounts "there is normative pressure to *expand* the range of situations in which it's proper to subject agents to norms of moralized praise/blame…" (McGeer 2015a: 2645). And she suggests, thus, that "agents are 'deserving' of praise and blame (i.e. they merit it) just in case they are the kinds of creatures that can be so scaffolded—i.e. they are responsive to the proximal moral-agency-fostering effects of our responsibility practices" (McGeer 2015a: 2647) One way of cashing out the relevant skill-based capacity in line with these considerations might be like the classical example of a skill, bike-riding. Once one learns how to ride a bike, one maintains the temporally-extended skill-based capacity to ride bikes even in circumstances in which one does not utilize that capacity. Unlike bike-riding, however, one might think that plausibly almost all human adults develop the skill-based capacity to be receptive to moralized blame. Thus, one might think that on this reading, almost all adult human beings are susceptible to blame *whenever* they do a bad action *regardless* of any mitigating circumstances.

So, one way of understanding McGeer here is to suggest that she claims that what grounds assessments of blameworthiness is that the agent maintains the temporally extended skill-based capacity to be receptive *in general* to moralized blame. But this leads to clearly absurd results. For example, one might think that according to McGeer's account it would be appropriate to blame Fallipe. Fallipe has been pushed out of a second story window. He lands on an unfortunate woman, crushing her. Is he blameworthy? One should think: clearly not. Even putting aside questions of desert, it's clear that blaming Fallipe will do nothing to structure his future agency. But notice that by McGeer's lights, we shouldn't appeal to any atemporal modal feature of Fallipe's agency to explain why he is not blameworthy. So we cannot simply say that Fallipe lacked the appropriate *control* over his behaviour as cashed out by his lack of reasons-responsiveness, and for that reason ought not to be held responsible.

However, one might claim that given that we are discussing responsibility as such, we're presuming a notion of agent-causal responsibility (Vargas 2022: 12-14). That is, given that we're asking the question of whether or not Fallipe is *responsible,* we're presuming that it is *his* action over which he may be liable to blame. Thus, McGeer might claim that given that Fallipe did not exercise any form of voluntary control over his falling out of a building, he is not blameworthy.

However, one prima facie worry for responding in this way is that on standard compatibilist accounts of moral responsibility, the distinction between mere causal responsibility and the voluntary control that would undergird moral responsibility is adjudicated precisely by means of those 'atemporal modal' accounts which McGeer herself rejects. However, McGeer may construct an analogous account; she may claim that agents are blameworthy to the extent that they perform a bad action and are scaffoldable with respect to those moral considerations *that they can be made responsive to in the future*. Cashing out scaffoldability in this way allows McGeer to excuse Fallipe; he has (plausibly) no control over whether he will be defenestrated in the future.

But this, too, seems to lead to problematic results, particularly in the context of omissions. For example, consider Snoozy. Snoozy often takes short naps during the day; despite being short naps, Snoozy typically falls into a deep insensitive slumber during them. During one such short nap on a park bench near a duck pond, a child falls into the duck pond and screams for help; Snoozy is unresponsive. Is Snoozy blameworthy? I think it's reasonable to take Snoozy to be morally in the clear: he was not responsive to moral reasons given that he was asleep, and he had no reason to think he would need to be responsive while he was asleep on the park bench.

While reasons-responsiveness accounts appear to have the resources to assess that Snoozy, and agents like him, are not blameworthy, McGeer's account appears to lack the resources to exculpate individuals in the context of omissions. After all, while Snoozy does not maintain reasons-responsiveness, he maintains (plausibly) the

temporally-extended skill-based capacity to be receptive to moralized praise and blame with respect to the consideration that he ought to help drowning children. And McGeer explicitly rejects appeal to the resources standard compatibilist accounts appeal to in order to adjudicate the responsibility of such agents, e.g., atemporal features of his agency and the situation, such as the fact that he is asleep and had no awareness that he ought to be awake.

One response that McGeer might make here is to bite the bullet and claim that Snoozy *is* blameworthy. If all that matters to us is that in blaming agents we serve to scaffold their moral responsibility, then, one might think, we could scaffold Snoozy's SMCRA by blaming him for not attempting to save the drowning child, *even though* he lacked the appropriate reasons-responsiveness to be held blameworthy on that account.

But it is important, I think, that even if we accept that blame generally has the appropriate proleptic effects on its targets that we do not too radically expand the set of targets licensed as being appropriate for blame on our account. If we do so, it seems likely to me that we would likely end up undercutting the proleptic force of blame in general. That is, if we regularly blame people for action which they do not take themselves to be responsible for, then, far from encouraging the uptake of the relevant norms, we are likely to foster defensiveness and anger and *undercut* the internalization of those norms.

Consider, for example, implicit bias. As Vargas (2017: 237-8) argues, "if we begin insisting on universal culpability for implicit biases, the risk is that we provoke widespread defensiveness and hostility. Defensiveness and hostility might slow the successful internalization of the relevant norms, and might even undercut the moral force of concern for bias, its effects, or even practices of moral blame." This seems to be supported by the empirical evidence; Vitriol and Moskowitz (2021: 2) point to evidence suggesting that blaming people for implicit bias "has the opposite of the desired effect of regulating bias, triggering instead a constellation of self-protective responses comprising resentment, denial, denigrating the source of the feedback, and trivializing the importance of the domain in which the feedback occurs." Instead, they claim, an appropriate uptake of norms surrounding implicit bias may be achieved when that feedback is such that agents do not take themselves to be *blameworthy* for having, and acting on the basis of, those biases.

So it looks as though McGeer likely ought to find some way to claim that Snoozy is not blameworthy, because it seems likely that blaming Snoozy and agents like him would undercut the proleptic force of blame in general. One way that McGeer could respond is to claim that Snoozy is not scaffoldable with respect to that *consideration* in that *kind of circumstance;* I think this is what we should say as instrumentalists, and this is why I ultimately suggest that reasons-responsiveness functions as a better grounding condition for instrumentalist accounts than influenceability.[6]

6 Note, too, that there are other reasons to be sceptical of influenceability on the basis of what it says about, e.g., curmudgeons who act rationally, but are entirely unreceptive to the influence of moralized blaming, and the aptness of practices such as blaming the dying or absent (Vargas 2015).

However, note that it's difficult to see how McGeer can respond in this way while maintaining her temporally-extended skill-based account. Once we begin appealing to atemporal modal features of, e.g., Snoozy's situation, it is difficult to see what the difference between McGeer's temporally extended skill-based account, and an atemporal modal account, such as a reasons-responsiveness account, is supposed to be. Either I am wrong in my interpretation, and Snoozy, and agents like him, should be blamed on McGeer's account (but this seems problematic), or I am right; if I am right, we're, at the very least, owed an explanation of how she can exculpate agents like Snoozy.

Further, if I am right, this leads to two problems for her. The first problem, as I noted, is that making this move would appear to mean abandoning her temporally-extended skill-based account. The further problem is that, were she to abandon her temporally extended skill-based account for an atemporal modal account this would, she thinks, ultimately undercut the proleptic force of blaming action given the Hard Problem, and the evidence from social psychology which suggests that there may be many more circumstances in which we may be circumstantially diminished in responsibility, and that exculpating individuals in such circumstances will not serve the proleptic desideratum.

4 The Hard Problem of Moral Responsibility vs. Reasons-Responsiveness

As we saw above, McGeer believes that the Hard Problem poses a grave problem for reasons-responsiveness accounts. However, Vargas' defence of his account as against the Hard Problem is very simple. As he (2015: 2677) puts it, "my own response is simple. *Them's the rules, and them's good rules.*" That is, I take it, if the Hard Problem motivates a theoretical concern with regard to the *desert* desideratum, this isn't a problem for Vargas. That such a concern might make his account seem inapt given the desert desideratum is, for Vargas, ultimately irrelevant. This is because what matters on his account is not desert, but instead the proleptic upshot of utilizing reasons-responsiveness for grounding blame attributions. When we blame apparently reasons-responsive people, Vargas is claiming, this will in general serve to cultivate their SMCRA, and *that's* what matters.

However, I think that McGeer could reasonably claim here that Vargas' response is too quick given the Situationist evidence that also partially motivates Vargas' account (see, e.g., Vargas 2013b). This Situationist evidence from social psychology appears as though it shows that agents' reasons-responsive capacities are fully manifest in far fewer situations than we might have pre-theoretically thought. According to this research[7], fine-grained situational features appear to have surprisingly strong influences on individual behaviour. For example, loud noises, being within a group of people, being asked to hurry, and being asked by a man in a white lab coat to shock someone to apparent

7 For discussion of how to interpret and respond to the evidence from social psychology, see Doris (2002; 2015), Vargas (2013b), and Waggoner et al. (2022).

insensitivity and death, all seem to radically diminish individuals' responsiveness to moral considerations in those circumstances.[8] The concern is that this evidence implies that there may be many more such situational features that are radically constraining of individuals' reasons-responsive capacities.[9] In this respect, one might take the Hard Problem to also constitute an empirical problem for Vargas given the *proleptic* desideratum.

Thus, one might worry that because the Situationist research suggests that there may be many more fine-grained situational features which are constraining of individuals' reasons-responsiveness capacities in those situations, it might appear as though it really is the case that in many—maybe even most, or all—circumstances in which an agent fails to act appropriately, that failure is ultimately down to a situational influence not under their control.

One can see that Vargas is very conscious of the possibility of this widespread lack of control in adjudicating responsibility. A substantial portion of his account consists in a scrupulous attempt to appropriately adjudicate agential control in light of this Situationist evidence. By his lights, alternative accounts of agential control maintain the fault of *atomism*. That is, they are atomistic—they assume that agential control can be adjudicated purely in terms of features internal to the agent themselves—but the evidence from Situationism suggests that agency is always, importantly, the product of agents *in* particular circumstances. Vargas (2013a: 3) thus advocates *circumstantialism about responsible agency*: the claim that agency is much more circumstantially fragmented than we might pre-theoretically think; otherwise apparently responsible agents may, in fact, not be responsible in many more circumstances than we might pre-theoretically suppose.

But the problem is that being so scrupulous about agential control might ultimately undercut his account's congruence with the proleptic desideratum. This is because McGeer (2019: 313) thinks that our SMCRA is "susceptible to development and decay—in that sense [it] is fragile" and that "it also depends, in very large part, on getting the right sort of feedback from the environment, where such feedback may be essentially social in nature."

This means that, according to McGeer, there is a prima facie conflict between the grounding of blame attributions on the basis of reasons-responsiveness with the proleptic goals of Vargas' ACM. This is because the empirical evidence suggests that we may, much more often than we might pre-theoretically think, be circumstantially constrained in our moral agency given fine-grained features of the situation in which we find ourselves, such as loud noises, being asked to hurry, etc.; were we to restrain the appropriate targets of blame in light of this empirical

8 See, respectively, Mathews and Canon (1975), Latane and Darley (1968), Darley and Batson (1973), and Milgram (1974).

9 One might object at this point: all of this research is decades old and subject to the concerns raised by the replication crisis in the social sciences; we shouldn't take it seriously. I have two responses to this concern: the first is that within the moral responsibility literature, people *do* take it seriously; see Vargas (2013b), Doris (2015), and McGeer (2015a). Secondly, despite some purportedly countervailing experimental results, the debate surrounding the evidence from social psychology is far from concluded, with several experiments apparently replicating successfully (see, e.g., Dolinski et al. [2017] and Fischer et al. [2011]).

evidence, we would be undercutting the proleptic efficacy of our blaming practices. Instead, we ought, McGeer believes, to *expand* the range of appropriate targets for emotive condemnation. For this reason, McGeer's scaffolding account is, she notes, "a substantially more modest requirement on responsible agency" (2019: 315) than Vargas', though I've argued that it is even *more* modest—problematically so—than she supposes.

Thus, the problem pointed to by the Hard Problem might appear to be more substantial than Vargas recognizes. It suggests that if we utilize Vargas' account of reasons-responsiveness, it may turn out to be the case that many fewer—and perhaps no—individuals are, in fact, blameworthy. And this, McGeer claims, will undercut the appropriateness of Vargas' proposed grounding of blame attributions given the proleptic desideratum. And given that both accounts are ultimately centrally grounded on the proleptic desideratum, this observation might appear to pose a big problem for Vargas' account.

That said, I do not think that the empirical reading of the Hard Problem is actually a substantial problem for proleptically-justified reasons-responsiveness accounts. To see why, consider Christian:

> Christian is a seminary student. He is asked to deliver a lecture on the parable of the Good Samaritan for a campus radio programme. However, he is also told that the programme has already started and he is already a few minutes late. He rushes to the building on campus in which the programme will be broadcast. On the way, he passes through a corridor in which an older man is slumped over groaning and coughing. He does not stop to help.

It looks as though Christian lacks responsiveness to the moral consideration, ironically, that he ought to help distressed individuals. The evidence appears to suggest that in such circumstances of high time-pressure, individuals like Christian may lack the appropriate responsiveness to moral considerations. This will straightforwardly be the case if such individuals lack the appropriate capacity to *detect* such agents as in need of help; this is what several individuals in scenarios identical to Christian's reported. (Darley and Batson 1973: 107-108; Vargas 2013b: 339-340).[10]

So, might this kind of evidence suggest a problematic diminishing of culpability on Vargas' reasons-responsiveness account? The worry that this evidence points to is that there may be many more such circumstances in which our reasons-responsiveness capacities are circumstantially diminished; this is what motivates Vargas' move away from atomism and towards circumstantialism regarding responsible agency.

A first point to note is that it seems unlikely that the influenceability condition will fare better than reasons-responsiveness in the scenario which Christian lacks the appropriate capacity to detect the relevant moral considerations in that circumstance. This is because it is unlikely that blaming Christian will serve to cultivate his

10 Alternatively, Christian may *detect* the relevant moral consideration, but struggle to *react* appropriately to it. In this scenario whether or not he maintains the appropriate control to be held responsible is more murky, though it's likely that most reasons-responsiveness accounts would hold Christian responsible; see Fischer and Ravizza 1998: ch. 3; McKenna 2005; Mele 2006; Vargas 2013b. Note that if reasons-responsiveness account *do* hold agents like Christian responsible, then the empirical reading of the Hard Problem I've suggested is defused; purportedly proleptically-effective blame would not be curtailed. Thus, I will set this reading of the scenario aside.

SMCRA in this circumstance. It is possible that blaming Christian will serve to encourage him to take more stock of similar moral considerations in similar sorts of circumstances in the future. However, if he simply lacks the appropriate capacity to detect moral considerations in those sorts of circumstances, then blaming him is likely to do nothing to rectify that. This may especially be the case given that Situationist research generally points to *surprising* situational influences on behaviour; if it turns out that the best self-regulatory strategy for Christian is to avoid similar situations in the future, then blaming him may do nothing to make him aware that he ought to go about self-regulating by avoiding similar situations in the future.

Additionally, reasons-responsiveness accounts have two resources that they can appeal to here. These are tracing (Vargas 2013: ch.5 and ch.9; Fischer and Ravizza 1998: ch.2), and the epistemic condition on responsible agency (Rudy-Hiller 2018). Reasons-responsiveness accounts that incorporate tracing claim that even if an agent is not reasons-responsive, if the reason that they are not reasons-responsive is some earlier decision that they made when they *were* reasons-responsive, then we can trace back their present responsibility to that earlier moment, and hold them responsible as a result. Whether or not we're justified in so holding them responsible for their action will, thus, turn on whether or not they should've known better—and this is adjudicated by means of the epistemic condition on moral responsibility.

The epistemic condition on responsible agency is typically thought to have two requirements (Vargas 2013a: ch. 7; Rudy-Hiller 2018). First, an agent must have reasonable foresight about the outcomes of their action. Second, they must maintain the capacity to recognize moral considerations. The relevance of that first requirement is obvious here: we can claim that as we become aware of empirical evidence suggesting apparent situational diminishing of control, we may become appropriate targets for reactive attitudes to the extent that we do not appropriately self-regulate to avoid the impact of those situational influences. The discovery of empirical evidence suggesting a circumstantial diminishing of responsibility, thus, *should* be initially undermining of responsibility in that circumstance.

However, any given empirical evidence suggesting an apparent diminishing of reasons-responsive capacities will only lead to a general under-cutting of justified blame on the basis of reasons-responsiveness accounts if we lack any appropriate self-regulatory control over that influence. And this, it seems to me, is the right result. We may become responsible for the influence of situational circumstances on our action to the extent that we ought to be aware of the relevant self-regulatory strategies to avoid their influence. But if we *do not have (and could not have had)* that awareness because there are no self-regulatory norms for responding to that situational feature, then it is difficult to see why we should be held responsible. After all, we do not seem to deserve to be blamed on the basis of such retrospective desert, and deploying reactive attitudes against us is also unlikely to appropriately shift our future behaviour.

One might think that this leads to a vindication of Vargas' account. However, there are a whole class of agents whom Vargas' ACM problematically exculpates given his construal of the control and epistemic conditions.

Recall that the epistemic condition consists of two requirements: a requirement to maintain the appropriate awareness of outcomes, and a requirement to recognize moral considerations. As Vargas (2013a: 202) notes, on his ACM account, "the requirement for recognizing moral considerations is counted as internal to the control requirement."

However, incorporating *this* epistemic requirement on responsible agency—as part of either the control or epistemic condition—is problematic for instrumentalist accounts, as Holroyd (2018) points out. Consider Calhoun's (1989: 398) example, whom I'm calling Harry:

> Imagine, for example, a man [, Harry] who always refers to women as 'girls' or 'ladies.' He […] is uncoerced into doing so and is in complete possession of normal adult reasoning faculties. Yet it seems he ought not to be blamed for linguistically infantilizing or patronizing women, for, from his point of view, one cannot reasonably expect him to see anything wrong with his actions. We may suppose that in his childhood, his father and mother referred to women as 'girls' or 'ladies.' He may also have come to understand that the former is flattering because it suggests youth and the latter simply polite. We may suppose that the people to whom he was exposed when he was growing up gave him examples only of this linguistic use and this understanding of its significance. From his point of view, it is natural to conclude that 'girl' is flattering rather than infantilizing and that 'lady' is polite rather than patronizing.

As Calhoun notes, it looks as though on standard compatibilist accounts, Harry is apparently not responsible for his infantilizing and patronizing language. He does not recognize the moral valence of his action—he believes that in so treating women, far from being offensive, he is being flattering or polite –and so, one might think, he is not responsible, if one presumes that being aware of the moral significance of one's actions is necessary for one to be morally responsible. But, as Calhoun also notes (405), "excusing excusable ignorance by withholding moral reproach inhibits the publicizing and adopting of new moral standards." But we presumably want to explain the inculcation of novel moral considerations on the basis of Vargas' ACM given the proleptic desideratum. So we're left with a puzzle for Vargas: it seems as though we should assess Harry as being blameworthy, given that doing so would serve the proleptic goals of the ACM. But it appears as though Harry lacks the appropriate control over his actions to be held responsible on the basis of Vargas' grounding condition; by Vargas' lights, we should exculpate him due to his lack of appropriate epistemic awareness of what the salient moral considerations are.

But this seems puzzling. If what we're trying to do is explain the *development* of our SMCRA, then it seems wrong to bake into that account the assumption that agents are only responsible if they maintain the capacity to detect and react to those moral considerations *which they already recognize as such* in that circumstance. Making this move might make sense given the desert desideratum[11], but seems odd for an account predicated on the proleptic

11 See, e.g., Wolf's (1989) account, according to which agents only maintain blameworthy desert to the extent that they are appropriately responsive to the 'true and the good'.

desideratum. After all, how do people like Harry *become responsive* to those moral considerations if they are never held to account for failing to be responsive to them?

This will be a general problem in the context of any novel moral considerations for Vargas' account. And, unlike in the case of Christian, it does not seem as though we can appeal to tracing: there does not seem to be any historical instance we can appeal to in which Harry was responsive to those considerations, and thus may be responsible for not appropriately self-regulating.

5 Minimal Grounding Emendation

In this section, I will sketch a suggested emendation of the grounding conditions of Vargas' ACM which resolves the above described epistemic requirement worry with Vargas' account, and also expands the set of reactive attitudes and associated responses rendered potentially appropriate by meeting the grounding conditions. I call this suggested emendation the Minimal Grounding Emendation (MGE). It is a *minimal* grounding in that it is less sensitive to epistemic desert than Vargas' ACM, and in that it gives a minimal prescription for the appropriate reactive attitudes to deploy in such cases of moral failure. In being such a minimal grounding, it, I think, better accommodates the proleptic desideratum and serves to better cultivate the development of SMCRA than the grounding conditions of Vargas' ACM.

My suggested emendation is centrally inspired by Fischer and Ravizza's prototypical reasons-responsiveness account.[12] I appropriate the claim that they make that an agent is an *apt target* for our *reactive attitudes* just in case they *maintain moderate reasons-responsiveness.* I will first explain moderate reasons-responsiveness, before explaining why it's important on my suggested emendation that the grounding condition of moderate reasons-responsiveness licences agents as being *apt targets* for our reactive attitudes, rather than being *blameworthy*.

What does it mean to maintain moderate reasons-responsiveness? Fischer and Ravizza's moderate reasons-responsiveness gives a distinct account of the epistemic agential features that render agents apt targets for our reactive attitudes as compared to the reasons-responsive grounding condition of Vargas' ACM. By Fischer and Ravizza's (1998: 72) lights, moderate reasons-responsiveness is maintained if an agent's actual and counterfactual responsiveness to moral considerations maintains an apparently understandable or rational pattern (as construed by a rational third-party observer) *given that agent's values, beliefs, and desires*. That is, what matters is that they are

12 Fischer and Ravizza also specify reasons-responsiveness as operating over sub-personal *mechanisms* as opposed to *agents.* I do not adopt this aspect of their account. For some reasons why a focus on mechanisms may be problematic, see McKenna (2013). For Fischer and Ravizza, the use of mechanisms is critical to exculpating manipulated agents. However, if we take the role of our reactive attitudes to be cultivating SMCRA (and not explaining desert), then I do not see why it should be appropriate to exculpate manipulated agents. This is because manipulated agents are, presumably, still responsive to blame, and other reactive attitudes, in cultivating their future agency, like Harry is. So I dismiss this theoretically problematic aspect of their account.

apparently reasons-responsive given their own values, not that they are so reasons-responsive given whatever the relevant moral considerations actually are independently of what they think they are. This is subtly, but critically, distinct from Vargas' reasons-responsiveness account.

Consider, again, the misogynistic Harry. Responding in a proleptically-appropriate fashion to Harry's misogyny will most likely require, I claim, *ignoring* the fact that he couldn't reasonably have known better what the salient moral considerations are. After all, if our reactive attitudes are the primary drivers of the development of SMCRA, as both McGeer and Vargas appear to suggest, then how else would we explain how *anyone* develops the appropriate sensitivity? Thus, on my suggested emendation, an agent is assessed as responsible just in case their action is apparently understandable and rational given their own values, beliefs, and desires.

If we assess Harry's responsiveness in terms of what *he* takes the relevant moral considerations to be given his beliefs, desires, and values then we can say that he *is* a responsible agent. Given that he is a responsible agent, reproaching him for his moral failure is, I think, appropriate, despite the fact that he couldn't reasonably have done otherwise given his epistemic history. Thus, I suggest entirely abandoning this aspect of the epistemic, or, for Vargas, control condition on responsible agency.[13] Whether or not an agent lacks awareness of the relevant moral considerations should have no bearing on whether or not they are responsible agents (though I suggest maintaining the other requirement of the epistemic condition on responsible agency in order to appropriately adjudicate responsibility for agents such as Christian, as discussed above).

As we've seen, adjudicating responsibility in terms of *control* appears to be critical given both the desert and proleptic desiderata; absent some assessment of control, McGeer's account may licence blame in too many cases of bad action. However, holding agents to be responsible depending upon whether or not they meet the second requirement of the epistemic condition—while critical for the desert desideratum—is, I'm arguing, superfluous for the proleptic desideratum. Agents who did not know what the relevant moral considerations were, and could not reasonably have known any better, should still be subject to our reactive attitudes in order for their SMCRA to develop, and thus for our grounding condition to appropriately meet the proleptic desideratum.

One might still worry that there is no guarantee that Harry will actually uptake the relevant SMCRA, having been blamed. However, as I've suggested, I'm only offering an emendation to Vargas' account. Thus, I see no reason to abandon ACM's two-tier structure, according to which the account's proleptic justification operates over our reactivity practices as a whole, as opposed to over individual acts. This, helpfully, allows us to say that blaming, or deploying other reactive attitudes against Henry may be appropriate even if Harry does not himself develop SMCRA in

13 Note that whether or not Fischer and Ravizza would themselves hold Harry responsible turns on how they would cash out the epistemic condition on responsible agency, but they do not themselves offer such a specification. See Fischer and Ravizza 1998: 26.

response to those reactive attitudes.

This is desirable for two reasons. First, this allows us to say that it is appropriate to hold Harry responsible even if he is curmudgeonly unreceptive to those reactive responses; excusing curmudgeons would likely not be in the best interests of the efficacy and stability of the practice as a whole (Vargas 2013a: 177-181). Second, it is likely that a substantial portion of our reactive attitudes may be *indirectly proleptically effective* on *third-parties* who witness those reactive attitudes (Shoemaker and Vargas 2021: 590; Moody and Nojoumian 2024). A large part of what explains our SMCRA is likely to be the product of our witnessing inappropriate behaviour which is marked as such by moralized reproach, such as blame.

Note, too, what this suggests about the apparent brittleness of proleptically-justified grounding conditions. Depending upon the empirical evidence, it may turn out to be the case that much of what explains the development of our SMCRA are such indirect effects on third parties. If this is the case more often than not, it may turn out to be the case that the account which best serves the proleptic desideratum is one which, like McGeer's, largely abandons an assessment of culpability in terms of desert. That is, it may turn out to be the case that blaming people regularly for the false imputation of bad action *is* proleptically effective (even though it is not on those individuals who are themselves blamed) because of the effects of such blame on third parties who *witness* that blame. I'm betting that this is not the case, but the point to emphasize is, again, that consideration of the actual empirical effects of our reactive practices is critical given the proleptic desideratum on instrumentalist accounts.

Consideration of those actual effects is also why I suggest moving to assessing agents' responsibility in terms of being *apt targets* for our reactive attitudes, as Fischer and Ravizza suggest. What does it mean to be an *apt target* for our reactive attitudes? Fischer and Ravizza's account gives an assessment of the control-condition portion of an assessment of *responsible* agency, which, by their lights, is not coextensive with *blameworthy* agency. On their account, an agent who meets the condition of moderate reasons-responsiveness is not necessarily, for that reason, appropriately held to be praise or blameworthy. They are instead merely an apt target for our reactive attitudes.

The point to note here is that if what we're concerned with is the proleptic effects of our reactions to bad action, why are we assuming that the only potentially effective reaction will be blame? As Holroyd (2018: 155) points out, "it is notable that when P. F. Strawson first introduced the reactive attitudes as constitutive of the social practice of holding responsible, his focus was not limited to praise and blame, but extended to a range of attitudes including resentment, gratitude, indignation, pride, and love. At least, then, it looks like a mistake to preclude, via conceptual fiat, the importance of other kinds of moral responses to moral failures." Both McGeer and Vargas appear to assume that the responses that matter are praise and blame. Vargas (2013a: 250) is explicit about his account being intended to justify,

specifically, the application of blame, and McGeer (2015b) takes blame to be critical because it serves the instrumental role of prompting a dialogue regarding the salient moral considerations.

But why focus on blame in particular? One might take the justification of blame to be the most difficult, and that this is why such accounts focus on blame (and praise, but only insofar as it is analogous to blame). This makes sense given the desert desideratum—blame is most difficult to justify given that it is the most unpleasant response to be subject to. But once we make the move to justifying our reactive attitudes on the basis of the proleptic desideratum, the focus on blame seems to me to no longer be appropriate; other responses may be just as, if not more effective, in cultivating our SMCRA (Holroyd 2018; Moody and Nojoumian 2024; Pereboom 2021).

In positing that agents are 'apt targets' for our responsibility practices, I suggest that Vargas should take no stand on which reactive attitudes and associated responses, in particular, are warranted in that situation. This is because I suspect that the empirical efficacy of our reactive attitudes will be radically circumstantially dependent. Whether or not we develop SMCRA in response to reactive attitudes will depend upon the situational circumstances (as in the Situationist evidence cited above), but also on our relationship to the person blaming us (Eisenberg et al. 2006: 668-669), what the relevant moral considerations are (Harinck and Van Kleef 2012), what attitudes are levied against us (Hoffman 2000), and how those attitudes are framed (Trevors et al. 2016), etc.. For example, having a parent point to the unfortunate effects of our bad action on our friend may be far more effective in structuring our SMCRA than being angrily yelled at for the same bad action by a stranger on the street. In this respect, I suggest that what should matter centrally for instrumentalist accounts is not so much circumstantialism regarding agential control, as Vargas suggests, but instead circumstantialism regarding the appropriate reactive attitudes to deploy depending upon the roles, circumstances, relationships, and history between the blamed and blamer.

6 Concluding Thoughts

I've shown why we should adopt reasons-responsiveness as a grounding condition, as opposed to influenceability, given the proleptic desideratum that centrally justifies instrumentalist accounts of responsible agency. However, I've argued that Vargas' ACM still problematically exculpates people given his construal of the control condition on responsible agency. I've put forward my own suggested emendation of Vargas' ACM which avoids this problem. However, I've emphasized that, unlike standard retrospectively-justified compatibilist accounts, instrumentalist accounts are brittle: they are very vulnerable to empirical evidence suggesting a potential lack of proleptic efficacy of our reactive attitudes. In light of this concern, I've suggested that a focus on blame may be misguided: it may turn out to be the case that other responses are more appropriate in cultivating moral agency. Thus, I

suggest that while an account of control is likely important in order to avoid the deployment of proleptically ineffective reactive responses, the more important question for instrumentalist accounts is *how* to respond in cases of moral failure. I've suggested that it is an open and, as yet, under-explored empirical question as to which reactive attitudes and responses are appropriate to deploy (if any) in which circumstances in cases of moral failure when an agent is assessed as responsible.

Acknowledgements

This paper could not have been written without many productive discussions between myself and Tillmann Vierkant and Dave Ward. It was additionally substantially improved through feedback given by (in alphabetical order): James Carey, Niklas Dahl, Anton Emilsson, Anneli Jefferson, Makan Nojoumian, and Sam Williamson. I would also like to acknowledge useful feedback received at the 2023 'Formalising Responsibility' conference at the University of Manchester, and the 2024 'Language of Moral Repair' conference at Lund University. In addition, I would like to thank two referees at Ergo for their feedback on an earlier draft of this paper, and three referees at Synthese for their very helpful feedback in making this version of the paper as good as it is. Thank you all!

Bibliography

Brink, David O., and Dana Kay Nelkin. 'The Nature and Significance of Blame'. In *The Oxford Handbook of Moral Psychology*, edited by Manuel Vargas and John M. Doris, 0. Oxford University Press, 2022. https://doi.org/10.1093/oxfordhb/9780198871712.013.10.

Darley, John M., and C. Daniel Batson. '"From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior'. *Journal of Personality and Social Psychology* 27, no. 1 (1973): 100–108. https://doi.org/10.1037/h0034449.

Doliński, Dariusz, Tomasz Grzyb, Michał Folwarczny, Patrycja Grzybała, Karolina Krzyszycha, Karolina Martynowska, and Jakub Trojanowski. 'Would You Deliver an Electric Shock in 2015? Obedience in the Experimental Paradigm Developed by Stanley Milgram in the 50 Years Following the Original Studies'. *Social Psychological & Personality Science* 8, no. 8 (2017): 927–33. https://doi.org/10.1177/1948550617693060.

Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139878364.

Doris, John M. *Talking to Our Selves: Reflection, Ignorance, and Agency*. First edition. Oxford: University Press, 2015.

Eisenberg, Nancy, Richard A. Fabes, and Tracy L. Spinrad. 'Prosocial Development'. In *Handbook of Child Psychology, Social, Emotional, and Personality Development: Social, Emotional, and Personality Development*, 646–718. Hoboken, UNITED STATES: John Wiley & Sons, Incorporated, 2006. http://ebookcentral.proquest.com/lib/ed/detail.action?docID=261363.

Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge Studies in Philosophy and Law. Cambridge University Press, 1998. https://doi.org/10.1017/CBO9780511814594.

Fischer, Peter, Joachim I. Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. 'The Bystander-Effect: A Meta-Analytic Review on Bystander

Intervention in Dangerous and Non-Dangerous Emergencies'. *Psychological Bulletin* 137, no. 4 (2011): 517–37. https://doi.org/10.1037/a0023304.

Frankfurt, Harry G. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68, no. 1 (1971): 5–20. https://doi.org/10.2307/2024717.

Fricker, Miranda. 2016. 'What's the Point of Blame? A Paradigm Based Explanation'. *Noûs* 50 (1): 165–83. https://doi.org/10.1111/nous.12067.

Harinck, Fieke, and Gerben A. Van Kleef. 'Be Hard on the Interests and Soft on the Values: Conflict Issue Moderates the Effects of Anger in Negotiations'. *British Journal of Social Psychology* 51, no. 4 (2012): 741–52. https://doi.org/10.1111/j.2044-8309.2011.02089.x.

Holroyd, Jules. 'Two Ways of Socializing Moral Responsibility: Circumstantialism versus Scaffolded-Responsiveness'. In *Social Dimensions of Moral Responsibility*, edited by Katrina Hutchison, Catriona Mackenzie, and Marina Oshana, 0. Oxford University Press, 2018. https://doi.org/10.1093/oso/9780190609610.003.0006.

Jefferson, Anneli. 'Instrumentalism about Moral Responsibility Revisited'. *The Philosophical Quarterly* 69, no. 276 (1 July 2019): 555–73. https://doi.org/10.1093/pq/pqy062.

Kohn, Alfie. *Unconditional Parenting*. Atria Books, 2006.

Latane, Bibb, and John M. Darley. 'Group Inhibition of Bystander Intervention in Emergencies'. *Journal of Personality and Social Psychology* 10, no. 3 (1968): 215–21. https://doi.org/10.1037/h0026570.

Maccoby, Eleanor E., and John A. Martin. "Socialization in the Context of the Family: Parent-Child Interaction." In *Handbook of Child Psychology*, 4th ed., vol. 4, edited by Paul H. Mussen. New York: Wiley, 1983.

Mathews, Kenneth E., and Lance K. Canon. 'Environmental Noise Level as a Determinant of Helping Behavior'. *Journal of Personality and Social Psychology* 32, no. 4 (1975): 571–77. https://doi.org/10.1037/0022-3514.32.4.571.

McGeer, Victoria. 'Building a Better Theory of Responsibility'. *Philosophical Studies* 172, no. 10 (2015a): 2635–49. https://doi.org/10.1007/s11098-015-0478-1.

McGeer, Victoria. 'Mind-Making Practices: The Social Infrastructure of Self-Knowing Agency and Responsibility'. *Philosophical Explorations* 18, no. 2 (June 2015b): 259–81. https://doi.org/10.1080/13869795.2015.1032331.

McGeer, Victoria. 'Scaffolding Agency: A Proleptic Account of the Reactive Attitudes'. *European Journal of Philosophy* 27, no. 2 (2019): 301–23. https://doi.org/10.1111/ejop.12408.

McKenna, Michael. 'Reasons-Responsiveness, Agents, and Mechanisms 1'. In *Oxford Studies in Agency and Responsibility*, Vol. 1. Oxford University Press, 2013. http://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199694853.001.0001/acprof-9780199694853-chapter-7.

Milgram, Stanley. *Obedience to Authority: An Experimental View*. New York: Harper & Row, 1974.

Moody, Kristoffer, and Makan Nojoumian. 2024. 'Blame: What Is It Good For?' *Philosophical Explorations*, 1–19. https://doi.org/10.1080/13869795.2024.2405523.

Nelkin, Dana Kay. *Making Sense of Freedom and Responsibility*. Oxford: University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199608560.001.0001.

Pereboom, Derk. *Living without Free Will*. Cambridge Studies in Philosophy. Cambridge: University Press, 2001.

Pereboom, Derk. 'Undivided Forward-Looking Moral Responsibility'. *The Monist* 104, no. 4 (1 October 2021): 484–97. https://doi.org/10.1093/monist/onab014.

McGeer, Victoria, and Philip Pettit. 'The Hard Problem of Responsibility'. In *Oxford Studies in Agency and Responsibility*. Oxford: University Press, 2015. https://doi.org/10.1093/acprof:oso/9780198744832.003.0009.

McKenna, Michael. 2005. 'Reasons Reactivity and Incompatibilist Intuitions'. *Philosophical Explorations* 8 (2): 131–43. https://doi.org/10.1080/13869790500091508.

Mele, Alfred R. 2006. 'Fischer and Ravizza on Moral Responsibility'. *The Journal of Ethics* 10 (3): 283–94.

Rudy-Hiller, Fernando. 'The Epistemic Condition for Moral Responsibility'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/fall2018/entries/moral-responsibility-epistemic/.

Sher, George. *Who Knew?: Responsibility Without Awareness*. USA: Oxford University Press, 2009.

Shoemaker, David, and Manuel Vargas. 'Moral Torch Fishing: A Signaling Theory of Blame'. *Noûs (Bloomington, Indiana)* 55, no. 3 (2021): 581–602. https://doi.org/10.1111/nous.12316.

Strawson, P.F. 'Freedom and Resentment'. In *Freedom and Resentment and Other Essays*, 1–28. London; New York: Routledge, 2008.

Trevors, Gregory J., Krista R. Muis, Reinhard Pekrun, Gale M. Sinatra, and Philip H. Winne. 'Identity and Epistemic Emotions During Knowledge Revision: A Potential Account for the Backfire Effect'. *Discourse Processes* 53, no. 5–6 (3 July 2016): 339–70. https://doi.org/10.1080/0163853X.2015.1136507.

Vargas, Manuel. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: University Press, Oxford University Press, Incorporated, 2013a. https://doi.org/10.1093/acprof:oso/9780199697540.001.0001.

Vargas, Manuel. 'Situationism and Moral Responsibility: Free Will in Fragments'. In *Decomposing the Will*, 400–416. Oxford University Press, 2013b. http://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199746996.001.0001/acprof-9780199746996-chapter-17.

Vargas, Manuel. 'Desert, Responsibility, and Justification: A Reply to Doris, McGeer, and Robinson'. *Philosophical Studies* 172, no. 10 (1 October 2015): 2659–78. https://doi.org/10.1007/s11098-015-0480-7.

Vargas, Manuel. 'Implicit Bias, Responsibility, and Moral Ecology'. In *Oxford Studies in Agency and Responsibility Volume 4*, 219–47. Oxford University Press, 2017. http://www.oxfordscholarship.com/view/10.1093/oso/9780198805601.001.0001/oso-9780198805601-chapter-12.

Vargas, Manuel. 'Constitutive Instrumentalism and the Fragility of Responsibility'. *The Monist* 104, no. 4 (1 October 2021): 427–42. https://doi.org/10.1093/monist/onab010.

Vargas, Manuel. 'Instrumentalist Approaches to Moral Responsibility'. In *Oxford Handbook on Moral Responsibility*, 2022.

Vitriol, Joseph A., and Gordon B. Moskowitz. 'Reducing Defensive Responding to Implicit Bias Feedback: On the Role of Perceived Moral Threat and Efficacy to Change'. *Journal of Experimental Social Psychology* 96 (1 September 2021): 104165. https://doi.org/10.1016/j.jesp.2021.104165.

Waggoner, Maria, John M. Doris, and Manuel Vargas. 'Situationism, Moral Improvement, and Moral Responsibility'. In *The Oxford Handbook of Moral Psychology*, edited by Manuel Vargas and John M. Doris, 0. Oxford University Press, 2022. https://doi.org/10.1093/oxfordhb/9780198871712.013.32.

Wolf, Susan. 'Sanity and the Metaphysics of Responsibility'. In *Responsibility, Character, and the Emotions*, 46–62. Cambridge University Press, 1988. https://doi.org/10.1017/CBO9780511625411.003.

Wolf, Susan. *Freedom within Reason*. Cary: Oxford University Press, 1990.