

Does Situationism Threaten Free Will and Moral Responsibility?¹

Michael McKenna
University of Arizona

Brandon Warmke
Bowling Green State University

The situationist movement in social psychology has caused a considerable stir in philosophy over the last fifteen years. Much of this was prompted by the work of the philosophers Owen Flanagan (1991), Gilbert Harman (1999) and John Doris (2002). Each contended that familiar philosophical assumptions about the role of character in the explanation of human action were not supported by the situationists' experimental results. Most of the ensuing philosophical controversy has focused upon issues related to moral psychology and ethical theory, especially virtue ethics. More recently, the influence of situationism has also given rise to further questions regarding free will and moral responsibility (e.g., Brink 2013; Ciorria 2013; Doris 2002; Mele and Shepherd 2013; Miller 2016; Nahmias 2007; Nelkin 2005; Talbert 2009; and Vargas 2013b). In this paper, we focus just upon these latter issues. Moreover, we focus primarily on reasons-responsive theories. There is cause for concern that a range of situationist findings are in tension with the sort of reasons-responsiveness putatively required for free will and moral responsibility. Here, we develop and defend a response to the alleged situationist threat to free will and moral responsibility that we call *pessimistic realism*. We conclude on an optimistic note, however, exploring the possibility of strengthening our agency in the face of situational influences.

1. Situationism & the Fundamental Attribution Error

We begin with a few preliminaries. An initial challenge to assessing the influence of the philosophical situationist literature is that there is no consensus about what situationism is.² An extreme version is that there is no such thing as character—at least as it has been historically understood (e.g., Harman 1999: 328). That is, there exists nothing answering to a feature of persons consisting of a familiar set of “global” traits, especially those usually associated with ethical assessment. Persons simply are not the sorts of animals that are reliably honest or

¹ For helpful comments, we would like to thank David Brink, John Doris, Ishtiyaque Haji, Alfred Mele, Christian Miller, Dana Nelkin, Dan Russell, Paul Russell, Carolina Sartorio, David Shoemaker, Hannah Tierney, Jason Turner, and Manuel Vargas. We would also like to thank Christian Miller for kindly inviting us to contribute to this volume, and an anonymous referee for *Journal of Moral Philosophy*. Michael McKenna would like to thank the John Templeton Foundation for resources that helped support this project. Support for Brandon Warmke's work was funded in part by a grant from the Templeton World Charity Foundation. The opinions expressed in this paper are the authors' own and do not necessarily reflect the views of either foundation.

² Here, we gloss just a few ways of understanding the situationist thesis in philosophy. For a careful catalog of the ways 'situationism' has been understood by philosophers and psychologists, see Miller (2014: ch. 4).

dishonest, courageous or cowardly, generous or stingy, and so forth across a wide range of situations. At best, we have “local” traits: in special situation-specific contexts, persons are reliably honest, courageous, and the like. Being honest at home with one’s family just has no reliable trait-based relation to honesty in one’s workplace, or amongst friends, for instance (see, e.g., Doris 2002).

A distinct formulation has it that situation—and not character—determines conduct; character is largely (if not entirely) out of the causal loop (see, e.g., Epstein 1979: 1099; Bowers 1973: 319). While there is (or may be) such a thing as character, situation does all or nearly all of the causal work. A less dramatic thesis is that situation plays a greater role in determining conduct than character (e.g., Doris 2002: 24-5; Ross and Nisbett 1991: 113). This is consistent both with the thesis that there is such a thing as character (though perhaps not as traditionally conceived) and that it plays a causal role in generating conduct. Still even less dramatic is the thesis that situation plays a greater role in determining behavior than is *typically assumed*. This last thesis is consistent with character playing an equal or greater role in producing behavior.³

In what follows, we will work with the formulation proposed by Dana Nelkin, which grants both the existence of traditional character traits and their causal role in producing action:

(S) Traditional personality or character traits like honesty, kindness, or cowardice play less of a role in predicting and explaining behavior than do particular situational factors. (2005: 182)⁴

Nelkin’s formulation is similar to the one we attributed to Doris (2002) and also to Ross and Nisbett (1991) in the preceding paragraph. However, one potential difference has to do with Nelkin’s commitment to the existence of traditional character traits, and to their playing a causal role in intentional action. It is not clear that the likes of Doris or Ross and Nisbett are committed to the existence of character as traditionally understood. Rather, what they seem to be committed to is, instead, the more revisionary thesis that there is such a thing as character, but not as traditionally understood. There exists a collection of local traits comprising a person’s character, and it is character *so understood* that plays a causal role in human conduct. On this point, for reasons we will set out below, we side with Nelkin.

According to its advocates, situationism is an unexpected discovery, one that stems from a pervasive mistake in understanding the causes of human behavior. This mistake is referred to as the Fundamental Attribution Error (FAE), which concerns:

³ This is another way of taking both Doris (2002) and Ross and Nisbett (1991). This gloss of situationism is consistent with the previous one for it may be the case both that situation plays a greater role than character *and* that situation plays a greater role than is typically thought. The latter thesis is the weaker one, however, for it does not imply that situation plays an equal or greater role in the production of behavior.

⁴ Nelkin’s (S) is fairly liberal in that it targets a certain claim about *both* personality traits and character traits (of the moral and non-moral varieties). For our purposes, discussion of traits should be taken to refer to moral character traits. Because our discussion concerns the threat of situational influence vis a vis free will and responsibility, however, we don’t think much hangs on this limitation.

People's inflated beliefs in the importance of personality traits and dispositions, together with their failure to recognize the importance of situational factors in affecting behavior. (Ross and Nisbett 1991: 4)

Supposing (S) is true and that (FAE) is in play, there is a natural way to explain (FAE) in folk understanding. In many contexts, third person observation takes situational features to be transparent. What needs explaining is something about the internal life of the agent, which is not so transparent (Ross and Nisbett 1991: 141). So character traits and motives are emphasized, and situational features are taken as given. This can mislead one into thinking that in a complete or comprehensive explanation, it is character that plays the greater role, not situation.

In any event, whether or not (S) is true, and whether (FAE) is a pervasive fallacy in folk explanation of human action, we disagree with the more dramatic views associated with situationism, such as the view that character as traditionally understood plays little or no role in the causes of action, or even more controversially that there just is no such thing as character as traditionally understood. The evidence for such theses is based upon studies showing that when situation varies, a person cannot be expected to act from the same character trait that she might act from in a different situation. But this is poor evidence for these strong views (e.g., see Kamtekar 2004; Miller 2014: ch. 4; Russell 2009; Sreenivasan 2002).

To explain, suppose it is claimed that honesty in the work place does not necessarily manifest itself in honest behavior in one's home life. Imagine, for instance, that Sheri is supremely honest with her boss and coworkers but that she lies incessantly to her spouse Pat. Now, is Sheri's lying to Pat evidence that she lacks a relatively stable, situationally invariant character trait of honesty outside of her work environment? At first it might seem so. Indeed, it can be taken as defeasible evidence. But that is all. Note that character traits are dispositions, and dispositions can be retained even when not manifested, or not manifested in typical ways. Suppose that Pat is abusive or manipulative with Sheri and so Sheri's fear of Pat or her worries about Pat's manipulative ways leads her to mask her inclination to honesty so that she suppresses her acting upon it in her relations with Pat. In this case, some *other* disposition of Sheri's, maybe one involving fear, is at odds with her disposition to honesty. Moreover, Sheri might feel terrible or anxious about lying to Pat, and this might give rise to other ways of thinking about and treating Pat that are the product of her honesty. In this case, her character trait of honesty *would* be manifesting (and not suppressed) but by means other than outward honest behavior, such as her feeling terrible and then acting in relevant ways as a result of her bad feelings.⁵

The salient point is that as psychologically complex creatures, it should not be surprising that in the mix of various character traits an adult person might have, different traits might work against each other such that one suppresses the other, or by manifesting in less transparent ways than by direct behavior linking trait to act-type (e.g., honesty to truth-telling). However it works out, and whatever character comes to as traditionally understood, there is not much reason to think that the very notion of character, or instead its playing a nontrivial causal role in human conduct, is impugned by the situationist literature. Regardless, this is consistent with Nelkin's (S).

⁵ Kamtekar (2004: 473), for example, developed such a strategy for replying to Doris. For discussion of this "competing virtues" response, see Miller (2014: 221-223).

2. Situationism and the Situationist Literature

We turn now to a compressed sketch of several situationist studies, offering only brief commentary at this point. For purposes of brevity we will rely on our readers' general familiarity with (most of) these studies, or absent this, their faith that we are reporting them adequately.⁶

2.1 Hartshorne and May

One of the earliest empirical studies suggesting a kind of situationist view of human psychology was Hartshorne and May's 1928 famous study of elementary and secondary school children. Students were observed in a broad range of classroom and non-classroom situations so as to test their honesty. Students' levels of honesty were tested along a number of different measures, including: (1) their willingness to *steal* small amounts of money when they believed they were unobserved; (2) their willingness to *cheat* on a test when getting caught appeared to them to be highly unlikely; and (3) their willingness to *lie* about their own dishonest behavior. The reported results were surprising. On the one hand, children who were likely to, say, steal in one situation were also highly likely to steal in similar situations (with correlations as high as 0.79). However, Hartshorne and May reported that just because a student was likely to steal, this did not mean that she was also likely to, say, cheat on a test. Indeed, the correlations between the cross-situational measures were low enough (0.23) to undermine the kind of predictive power one might expect if there was a robust, cross-situational trait such as honesty. The upshot, it might be thought, is that whether people manifest honesty-like behavior is dependent on the kind of situation one is presented with. Whether one is honest largely depends on what kind of situation one finds oneself in.⁷

2.2 Milgram

Perhaps the most cited situationist studies are those that Stanley Milgram performed in the 1960s. We will focus on just one of them. Subjects were presented with what they were told was an experiment to test the efficacy of a particular teaching strategy: that learning is most effective when incorrect responses are punished. To test this theory, subjects were asked to administer shocks whenever another subject (in reality, a confederate of the study) gave the wrong answer to a question. The confederate was placed in an adjacent room, leaving the subject and researcher (a second confederate) alone in a room. The subject was seated before a machine that would "deliver the shocks" when the subject flips a switch. The rules: beginning with 15V, and in increasing increments, each successive incorrect answer was to be greeted with a shock, which were labeled and displayed on the machine as: slight, moderate, strong, very strong, intense, extreme intensity, and danger: severe shock-XXX. The experiment was designed to monitor how compliant the subject would be when she believed herself to be shocking someone (don't worry, no actual shocks were given). Indeed, pre-recorded reactions to the shocks, heard over a speaker ranged from a surprised "Ouch!" to screaming and banging on the wall. At 330V, the

⁶ We borrow a bit here from Rodgers and Warmke (2015).

⁷ For more on the Hartshorne and May study, including a number of critical discussions, see Burton (1963).

confederate ceased to respond. 65% (26 of 40) of subjects administered shocks through to 450V, long after the confederate stopped responding.

Here again, it appears that whether one engages in a certain kind of morally significant behavior crucially depends on one's situation. From reading the Milgram studies, it would be natural to draw the conclusion that while we would never do something we believed would probably kill a stranger in our everyday lives, many people just like us are willing to do so in a stressful environment when given persistent directions by an authority figure in the service of science. Is this all it takes to get people to do something they believe to be extremely harmful?

2.3 Zimbardo

Philip Zimbardo's 1971 "Stanford Prison" experiment is one of the most disheartening pieces of the situationist literature, not least because of how, by his own admission, Zimbardo himself got caught up in the scenario, which split undergraduate volunteers into two groups. Those in the convict group were "arrested" on campus before being taken to the prison, which was staffed by the remaining volunteers, who served as prison guards. Zimbardo himself served as the warden and gave one rule to the guards: no physical violence during the two-week period. None was needed. In just a few days, mattresses, food, and toilets became luxuries; prisoners were stripped and subjected to sexual humiliation. Zimbardo did not call off the experiment until an outsider suggested it had gone too far. He later admitted that he was to blame for being weak and "unable to handle" the experiment.⁸ The experiment ended after six days, when numerous participants complained of despair and depression. Again, situation seems to have radically altered people's behavior, turning otherwise normal college students into authoritarian monsters, even though they knew this was all role-play.

2.4 Isen and Reeve

In a 2005 study, Isen and Reeve investigated the influence of positive affect on task choice and completion of work. In one experiment, subjects who were offered a \$2 box of chocolates or candies (their choice!) for completed work tasks (such as circling a string of alphabetically ordered letters amid a string of random letters) more quickly, and as accurately, as those who received no gift. After completing their tasks, both groups were given time to work on subjectively desirable tasks. The subjects who had been given the gift also rated that desirable task as even more enjoyable than the group that did not receive the candies or chocolate. This study builds on an earlier and well-known experiment by Isen and Levin in which people who found a dime in a pay phone were said to be more likely to help others. While there is some debate about whether that study has been replicated successfully,⁹ the Isen and Reeve studies suggest that humble, unexpected, but desirable gifts can cause people to work more efficiently and enjoy what they do. But if correct, does this mean that how hard and efficiently we work might be much less a function of how industrious and hard-working we are, but instead whether we find ourselves in situations that put us in a good mood?

2.5 The Upshot

⁸ See <http://prisonexp.org>.

⁹ See Miller (2013: 63ff.) for discussion.

Note that these experiments (and the literature that has developed around them¹⁰) offer—at least on their face—some presumptive evidence in favor of (S). And to the extent that their results are surprising and at odds with expected behaviors, they also appear to confirm (FAE). A natural expectation in, for instance, the Stanford Prison Experiment is that most college kids would not be disposed to treat their peers so poorly. It is natural, we think, to impute to them traits such as decency, minimal kindness, and compassion for others. However, (S) helps explain why, whatever character traits one might be inclined to project onto this population qua college students, situational forces can “wash over” these traits and yield surprising behavior. (FAE) in turn, helps explain why the results are after all taken to be surprising. Similar remarks apply to each study mentioned.

We now wish to set aside whether or not these studies and numerous others really do support (S) and (FAE). Our focus here is not on character *per se*, and so is not directly focused on (S), but on free agency and moral responsibility. Thus, even if careful work on the nature of character can be used to discredit (S) and to diminish the relevance of (FAE), these studies might nevertheless provide independent reasons to question familiar attributions of freedom and responsibility. Indeed, note that on many theories of freedom and responsibility, it is possible for a person to acquire her character through means that are freedom-and-responsibility-defeating in the sense that action flowing from a character so formed can be regarded as unfree because of how the character giving rise to it was acquired (e.g., Haji 1998; Mele 1995; Wolf 1987; and Watson 1987). So even if these studies do not support (S), they still might pose a threat to freedom and responsibility. Thus, we follow Nelkin (2005: 191) in distinguishing between *situationism* and the *situationist literature*. Even if situationism proper—say, in the form of (S)—is to be rejected, do the preceding studies alleged to support (S) still raise trouble for free will and moral responsibility?

3. *Reasons-responsive Theories of Freedom and Responsibility*

We ask for the reader to set aside for the moment our discussion of situationism and the situationist literature. We now turn to the topic of moral responsibility. We will then circle back to discuss the putative threat to freedom and moral responsibility posed by the situationist literature.

Numerous philosophers working on the topics of free will and moral responsibility have attempted to account for freedom in terms of an agent’s responsiveness to reasons.¹¹ An agent who acts freely is responsive to *variation* in a suitable spectrum of reasons. That is, to be free and morally responsible requires a certain kind of ability to respond to reasons in certain kinds of

¹⁰ See, for example, Asch (1951), Blass (1966), Darley (1995), Darley and Batson (1973), Doris and Murphy (2007), Elms (1995), Goldhagen (1996), Haney, Bank, and Zimbardo (1973), Haney and Zimbardo (1998), Latane and Darley (1970), Lusky (1995), Meeus and Raaijmakers (1995), Miller (1995), Modigliani and Rocha (1995), Newcomb (1929), Rocha and Modigliani (1995), Rosenthal (1999), Sears (1963), and Zimbardo (2007).

¹¹ See, for example, Brink and Nelkin (2013); Dennett (1984); Fischer and Ravizza (1998); Gert and Duggan (1979); Haji (1998); McKenna (2013); Nelkin (2011); Sartorio (2016); Vargas (2013a); and Wolf (1990).

ways. Because many theorists understand free will in terms of the control condition(s) necessary for moral responsibility, reasons-responsive theories are also frequently cast in terms of theories of the control condition for moral responsibility. This is how we will proceed.¹²

For present purposes we can gloss over important differences regarding how a reasons-responsive theory is best advanced. Four brief points will suffice for present purposes:

First, reasons-responsive theories have a *modal dimension*. Let us explain. An exercise of freedom involving moral responsibility involves acting from resources whereby an agent is responsive or sensitive to reasons. The modal dimension concerns the relevant sense of responsiveness or sensitivity. In many cases an agent's acting from reasons-responsive resources will involve her *actually* responding to pertinent reasons. If, for instance, she is blameworthy and acts wrongly from a motive of selfishness, this might involve her responding to a reason that such and such course of action will increase her wealth. Nevertheless, it is not a requirement of an agent being reasons-responsive in acting as she does that she *in fact* respond to any reasons at all. She might be acting from pure impulse or brute desire.¹³ What matters is that the resources from which she acts support dispositions allowing for the prospect of responding to reasons. In this way a free agent will be *able* to respond to reasons even if when she acts she does not actually do so.¹⁴ Crucially, on most reasons-responsive theories, for an agent to be morally responsible for what she does she must, in some sense of 'able', be able to respond to specifically *moral* reasons, regardless of whether she actually does so respond (e.g., see Fischer and Ravizza 1998; Nelkin 2011; and Wolf 1990). An agent *incapable* of responding to moral reasons is not free to act in ways that would make her morally responsible for how she acts.¹⁵

Because of this modal or dispositional aspect of reasons-responsiveness, an agent's being reasons-responsive supports the truth of counterfactuals regarding how an agent would respond in a range of nonactual but possible scenarios that are relevantly like the actual scenarios in which she finds herself. Suppose an agent acts on impulse when she responds hurtfully to a friend. She might nevertheless be reasons-responsive if there is a range of similar conditions in which, were different reasons "present", she would alter her conduct and not respond hurtfully. For instance, if she saw the friend was injured or depressed at the time, or if she was prompted by something in her environment to consider the likely effects of a hurtful remark, she would respond suitably.

Second, responsiveness to reasons, so understood, comes in *degrees*. A person rendered unfree and not responsible for an act of compulsive handwashing might still be responsive to some range of

¹² We acknowledge, of course, that there are other theories of freedom that are designed to account for the control condition on moral responsibility, for example Doris (2002). The question of whether and to what extent the situationist literature challenges those theories we will leave for others. Here, we shall only be concerned to explore whether and to what extent the situationist literature challenges reasons-responsive theories of freedom and moral responsibility.

¹³ We're aware that on certain theories of action all actions involve reasons as causes. We wish to make no commitments about this here. Of course, on such a view our current claim would need modifying. We'll not develop that here.

¹⁴ Accommodations have to be made for actual-sequence theorists who contend that the ability to do otherwise is not required for such theories (e.g., Fischer and Ravizza 1998; McKenna 2013; and Sartorio 2016). We'll not pursue this here, but for a brief explanation see McKenna (forthcoming).

¹⁵ For a dissenting opinion, see Talbert (2012).

reasons while being unresponsive to many good reasons for not handwashing. If, for instance, she were told not to wash her hands while being threatened at gun point, she might respond to this reason by refraining from washing. Likewise, a robustly free agent, responsive to many good reasons, might not be responsive to *all* good reasons bearing on her context of action. For example, a person who is praiseworthy for giving to a charity for starving children might be moved by many equally weighty moral reasons to give to different kinds of charities. And yet she might be blind to some similarly weighty moral reasons. Perhaps she doesn't see helping victims of AIDS as a similarly weighty moral reason and so wouldn't give to such charities. That she would not respond to all good (and sufficient) reasons to give to a worthy charity does not mean that she fails to act freely when she actually gives to a charity for starving children. What is required, then, on a reasons-responsive account of freedom is the specification of a *spectrum* of responsiveness to a sufficiently rich range of reasons. Doing so establishes that the agent is a sane, morally competent person. Fischer and Ravizza (1998) offer a specific proposal cast in terms of *moderate reasons-responsiveness*, MRR. For our purposes, we will deploy the label, MRR, without setting out the details of their view—although for the most part we are largely in agreement with their carefully cast proposal.¹⁶

Third, a reasons-responsive theory like MRR has both a *cognitive* and a *volitional* dimension (Brink and Nelkin 2013; Fischer and Ravizza 1998; and McKenna 2013). One way to flesh this out is to say, as Fischer and Ravizza (1998) put it, that MRR has a *receptivity* component and a *reactivity* component. An agent must first be able to recognize what good reasons there are and assess them for whether they are sufficient for pursuing a course of action. This is the cognitive dimension whereby reasons are “received.” But second, she must also be able to react appropriately so as to guide her own actions in light of her recognition of the good reasons that there are. This is the volitional dimension whereby an agent reacts to the reasons after she receives them.¹⁷

One reason to say that a theory like MRR has both a receptivity and a reactivity component is that distinguishing them helps to explain a range of excuses and exemptions. When an agent is, for instance, tied up or instead paralyzed, she'll not be able to translate some good reasons she would accept into action. In the case of compulsive handwashing, for instance, an agent might recognize that she has good reason not to wash her hands, but she might not be able to be reactive to such reasons. Her “reactivity” component is impaired, which explains our inclination to excuse such behavior. In other cases, an agent instead might be impeded from recognizing what good reasons there are, as perhaps happens when operating in poor lighting conditions or suffering from a delusional disorder. Here, too, we would be inclined to excuse, but in this case, it

¹⁶ For some reservations and qualifications, see McKenna (2013).

¹⁷ These two dimensions of MRR bring into relief a common distinction between two different ways of thinking about reasons: as either “normative” or “motivating.” As we understand MRR, a full theory of reasons-responsiveness will require adverting to both notions. First, insofar as an agent has an ability to receive or recognize something as a reason, she must have the ability to see a reason as a consideration in favor of (or as counting against), say, a certain course of action. She must be able to recognize a reason as having normative force. But second, insofar as an agent has an ability to respond or react to a reason, she must have the agential resources to translate her recognition of that reason into action. Her recognition of a reason, therefore, must have motivational force for her.

would be because we would judge the agent to have lacked the ability to recognize or receive the relevant good reasons.

Fourth we wish to point out that reasons-responsive theories like MRR are neutral between compatibilist and incompatibilist theories of free will and moral responsibility. It is true that MRR is typically deployed in the service of defending compatibilism, which can be understood as the thesis that free will and moral responsibility are compatible with the truth of determinism.¹⁸ But there is no reason why a view like MRR can't be deployed by incompatibilists who defend libertarianism about free will and moral responsibility. On such a view, free will and moral responsibility require reasons-responsiveness as well as a further condition: the falsity of determinism. In any event, the apparent threats to freedom and responsibility posed by situationist findings pose just as much of a problem for libertarians as for compatibilists. For even if libertarianism is true, were we to find out that situations frequently short-circuit an agent's functioning in a suitably reasons-responsive manner, then this would threaten freedom even if, when that sort of freedom operates unimpeded, it does so by satisfying the metaphysically more demanding conditions libertarians lay down.

4. How the Situationist Literature Appears to Threaten Reasons-Responsiveness

With the preceding sketch of MRR in hand, let us now return to the situationist literature and draw a bead on the threat that the empirical evidence might pose to freedom and responsibility. For starters, consider the way Dana Nelkin puts the threat:

One way of seeing the situationist cases—or at least some of them—as troubling is this: simply put, the subjects seem to be acting for bad reasons, or at least not acting for good reasons, and they seem stuck doing so. (2005: 199)

Note Nelkin's words "and they seem stuck doing so." They are crucial. This is because on a reasons-responsive theory like MRR, as our preceding discussion should have made clear, it is not evidence of an agent's lack of freedom that she simply *does not* act on any reasons, or that she does not act on good reasons. If so, we'd be forced to conclude that all reckless or immoral behavior is unfree and so excused. It is, rather, that the cases seem to suggest that the agents were stuck, and in this way, their freedom to be *able* to act from good reasons was impeded. Indeed, Nelkin re-emphasizes her point when she remarks that:

...the way in which the subjects seem to proceed raises a question about whether they *can* act for good reasons—in some important sense of "can." (201)

The absolutely crucial point is that the threat is cast in terms of the undermining of an ability; merely failing to exercise it in a certain way won't do—or so it seems. (We'll return to this point below.)

¹⁸ For present purposes, determinism can be understood as the thesis that at any time, given the past and the laws of nature, only one future is physically possible (van Inwagen 1983: 3).

We now call attention to a further worry not fully captured by Nelkin’s way of expressing the source of her concern. She casts the challenge in terms of whether agents are able to act for good reasons, and this flows from her particular way of defending a reasons-responsive theory (see Nelkin (2011) and also Wolf (1990) for a similar view). It is natural to associate Nelkin’s concern with an ability to act for *moral* reasons, a condition on an MRR view we noted above. But an even more pervasive worry is that, morally salient reasons aside, some agents in some of these situations might simply be rendered unable to act from a wider of range reasons as well, even including some not-so-good reasons. Some of the studies seem to suggest that reasons-responsive resources are not really what’s causally in play in accounting for the behavior. For instance, consider another famous study done by Darley and Batson (1973). In one part of the study, seminary students were asked to give a talk on the parable of the Good Samaritan, whereupon they were then to walk across campus to deliver it. In their path, they encountered a man slumped in an alleyway (a confederate). Although overall 40% offered to help, only 10% offered to help after they had been told that they should *hurry* across campus to give their talk.¹⁹

In such a case, there is some concern that increases in a person’s sense of urgency can prevent them from recognizing numerous kinds of reasons, even bad ones. Suppose, for instance, some of the seminarians were malicious pickpockets who, had they noticed a person in distress in a doorway, might have seen it as a reason to rob the person. Now just imagine that the way the priming of urgency works is that when a situation seems very urgent, at least some people are shielded from the resources needed to recognize much of their environment. If so, for certain domains, their reasons-responsiveness would be, in a sense, temporarily closed down or running “off line.” The point here is that situational features might not only undermine one’s ability to recognize good reasons (as Nelkin notes), but that due to one’s situation, agents might also get “stuck” with a more general inability to recognize various and sundry kinds of reasons, morally good or bad.²⁰ If certain kinds of situations tended to undermine this general sort of ability, then this would be cause for concern: it may be that we are free and morally responsible much less often than is typically supposed.

Situations might undermine a more or less general capacity to respond to reasons in at least two ways. First, recall from the previous section that MRR has a cognitive component—what Fischer and Ravizza call receptivity. One way for situations to undermine a more or less general capacity to respond to reasons is therefore in virtue of compromising the receptivity aspect of MRR. Situations may sometimes prevent an agent from being able to see or recognize something as a reason (good or bad).²¹ Alternatively, situations could incapacitate a more or less general ability to respond to reasons by undermining the volitional component—what Fischer and Ravizza call

¹⁹ For fuller discussion see Miller (2014: 98ff.)

²⁰ To be clear, we do not claim either that the empirical studies we have canvassed or the rest of the relevant present empirical literature outright licenses or requires the conclusion that these situational features undermine MRR in this way. We do think, however, that this conclusion is a natural and reasonable worry that one might have about the direction to which the empirical studies point. The reader may then understand our strategy here as one of showing just one way of allaying this worry, at least insofar as present state of the empirical literature is concerned. We thank Christian Miller for encouraging us to clarify this point.

²¹ This account of the (apparent) situationist threat is prominently highlighted in Nahmias (2007), although he does not explicitly express the challenge in terms of threats to reason-responsiveness.

reactivity. Here, reasons-responsiveness would be undermined because a situation has the effect of eliminating or severely impairing the agent's ability to act on a reason.

Naturally, there are other ways that situational features might in piecemeal fashion take "off-line" one's ability to respond to reasons. Perhaps situational features undermine *both* components of MRR. Or perhaps situational features somehow cause interference *between* these two components, such that the cognitive and volitional systems are, as it were, operating fine on their own but not in tandem. And of course, the situationist challenge to free will and moral responsibility need not be thought of as always undermining MRR in the same way. Some situational features might undermine receptivity; others might undermine reactivity. At any rate, for now we shall simplify matters and speak of the situationist challenge as the claim that situational features eliminate or instead significantly degrade or impair one's ability to respond to reasons in virtue of undermining either receptivity or reactivity.

Before proceeding, we pause to qualify the pertinent challenge. Note that just above we used the language "situational features eliminate or instead significantly degrade or impair one's ability." This needs clarifying. Obviously, the cleanest way to express the challenge is by restricting it to *eliminating* an ability altogether. This is in keeping with Nelkin's own way of formulating the worry in terms of an agent's being stuck in such a way that she *cannot* act for good reasons (2005: 199-201). But in our estimation, this would be too restrictive to give voice to a credible version of the situationist challenge we wish to explore. Since capacities and abilities come in degrees, they can be degraded significantly enough without being eliminated to count as fully exculpatory—and not just count as mitigating factors. Example: Joe, a professional basketball player, is an excellent free throw shooter who can hit roughly 80% of his shots. If he is tied up, his here-and-now specific ability is eliminated. He can hit exactly 0%. But suppose he can hit 40% with his left hand tied behind his back. Is his ability completely eliminated? It seems not. But is it degraded? Yes. Enough to count as fully exculpatory rather than mitigating (because it is freedom-defeating rather than freedom-diminishing)? Maybe not. But now place a deep cut in his right arm and pour salt water in his eyes so his success rate plummets to 3%. At this point, we are prepared to say that Joe's ability is retained but is so degraded that it is exculpatory (not in terms of moral blame, but perhaps in terms of "sports blame"). Likewise, we want to say, it is enough of a threat to freedom and responsibility if a situation significantly degrades an agent's reasons-responsive capacities or abilities. It need not completely eliminate them. Thus, in what follows we will freely write in terms of either eliminating or degrading these capacities or abilities.²²

To illustrate the challenge, recall the situationist studies set out above. To the extent that situation seems to have undermined the responsibility of the agents involved, it is plausible to think that it did so by compromising an ability to be responsive to reasons—by either eliminating it altogether or severely compromising it. Consider, for example, the subjects in the Stanford Prison Experiment. The default presumption, one that relies upon the powerful role of character,

²² We are indebted to Al Mele for his excellent comments here, as well as the example of a free throw shooter. Note that Mele advises a more cautious approach by separating cases of full elimination from cases of extreme forms of degradation of an ability. He is probably correct about this, however, it also seems likely that the best that most of the situationist studies could ever really do to generate the threat is degrade and not completely eliminate. So, simply for purposes of exposition, we'll make do with the less cautious formulation proposed here.

is that the vast majority of the subjects in this study who served as “guards” were generally good people who would never dream of humiliating and subjecting others to degrading treatment in their day to day lives. And yet in the limited confines of this experiment they did. One way of explaining this striking behavior is to hold that the subjects’ ability to respond to good reasons was severely impaired, and the environment produced the conditions that yielded this result. Perhaps due to the nature of the physical setting, the imposing authority figures, and the allure of imaginative play, the “guards” were simply unable to see certain reasons for or against certain ways to treating the “inmates.” They became unreceptive to an important range of reasons. Alternatively, these situational features might have severely impaired their ability to react to reasons: they saw that they should not degrade their charges but they were unable to translate those reasons into action.²³

If this is all that the situationist threat amounted to it would surely be cause for some concern. But would the threat be that serious? After all, how often do we find ourselves in faux prisons?²⁴ And yet it is not difficult to see that the situationist threat might very well generalize to all kinds of situations. If people are harder workers in large part because they receive lots of positive mood-enhancers throughout their lives, might they be less morally responsible for their hard work? Are those who work less hard also less morally responsible for their productivity because they were faced with very few mood-enhancing situations? And how deep might the rabbit hole go? If situational features of one sort or another are constantly bombarding our reasons-responsive capabilities and either seriously impairing them or taking them temporarily off-line, for how much of our behavior are we really moral responsible?

5. *A Plausible Defense in the Face of the Situationist Literature*

Recall from Section 1 Nelkin’s way of construing the situationist thesis:

(S) Traditional personality or character traits like honesty, kindness, or cowardice play less of a role in predicting and explaining behavior than do particular situational factors. (2005: 182)

Nelkin does not endorse the skeptical thesis at issue; she only offers it as a diagnostic proposal. On her view, *if* (S), or more precisely the situationist literature in support of (S), poses a threat to

²³ An anonymous referee has keenly noted that, given the set up in the experiment, an explanation in terms of the subjects’ severely impaired ability to respond to good reason converges (or nearly so) on one which the environment causes people to respond to bad reasons. As this referee notes, in the original paper reporting the experiment, the authors described the environment as “intrinsically pathological” (Haney, Banks, and Zimbardo 1973: 90). And in *The Lucifer Effect*, Zimbardo contends that “a general norm of permissiveness prevailed [at Abu Ghraib] that created a sense that the guards could do pretty much whatever they felt like doing because they were not personally accountable and could get away with anything because no one was watching” (2007: 368). The point here is that Zimbardo’s explanation highlights the environment as the root source of any pathology that would count as ability or capacity impairing.

²⁴ Though when we do find ourselves in such distressing situations, the outcomes are often not pretty. See for example, discussions of how contexts of war deleteriously affect behavior: Doris and Murphy (2007), Talbert and Wolfendale (2015).

freedom and responsibility, plausibly it is because it undermines reasons-responsiveness. At this juncture, we wish to note a way of resisting this skeptical threat as Nelkin has cast it, one with which we are largely in agreement, and, we suspect, Nelkin is too. The seeds of this strategy were already planted in our presentation of the threat. It is not enough to show that situational features play a substantial causal role in an agent's actual conduct, since her being reasons-responsive is consistent with her not actually responding to pertinent reasons. What's required to generate the skeptical threat, as Nelkin's careful wording made clear, has to do with a situation somehow corrupting an ability or capacity to respond to reasons. And actual behavior alone as described in the various situations from the literature is consistent with *retention* of the pertinent abilities or capacities. This provides the basis for a cautious way of resisting the skeptical challenge. Why not, for instance, conclude that in Zimbardo's Stanford Prison Experiment, many of the students playing the role of guards were culpable and blameworthy for their cruel treatment? The same goes for those who were so servile and cowardly.

David Brink (2013), who defends his own brand of reasons-responsive theory, develops the preceding point by building upon the following observation:

It is important to frame this approach to responsibility in terms of normative competence and the possession of these capacities for reasons-responsiveness. In particular, responsibility must be predicated on the possession, rather than the use, of such capacities. We do excuse for lack of competence. We do not excuse for failures to exercise these capacities properly. Provided the agent had the relevant cognitive and volitional capacities, we do not excuse the weak-willed or the willful wrongdoer for failing to recognize or respond appropriately to reasons. (131-2)

Note Brink's specification of cognitive and volitional capacities. While it might be true that situational features make a crucial causal contribution to *how* a pertinent capacity is exercised, we cannot take this as a basis for exculpation. What is required is that the contribution causally undermines or bypasses the pertinent cognitive and volitional (receptive and reactive) capacities. The key to Brink's resistance, then, in defense of a reasons-responsive theory, is the distinction between a *performance* error and a *competence* error (140). Very roughly, a psychological performance error is the result of the malfunctioning of what we might call a psychological mechanism. When there is interference or foul play—a "gumming up" of the works—the mechanism issues in a performance error. Consider, for example, a type of linguistic performance error called a spoonerism. Suppose, as you walk into the dean's office to discuss a very delicate matter, you ask the dean's assistant, "Is the bean dizzy?" Naturally, you meant to say, "Is the dean busy?" But just because you committed a spoonerism, this does not mean that you did not understand or know how to speak the correct English words. In fact, you have that competence. Yet you just had a slip of the tongue, perhaps due to some anxiety about meeting with a superior. Your situation somehow impeded the proper deployment of your linguistic competence.

One way to think of the individuals in, say, the Milgram experiments, then, is as having committed a performance error in the deployment of a normative competence, cashed out in terms of reasons-responsiveness. The thing to note here, however, is that a performance error is in itself not necessarily exculpating, only a lack of competence is, where the latter is accounted for in terms of reasons-responsive capacities. Brink writes:

[M]ost situational findings describe patterns, albeit surprising patterns, in behavior. They do not demonstrate incapacity. For instance, in the Milgram experiments we may be surprised to find that two-thirds of subjects fail to do the compassionate thing because of a desire, perhaps sub-conscious, to conform to the firm requests of the experimenter. But while many people did in fact act surprisingly badly, there's nothing in the situationist gloss in the Milgram experiments to suggest that subjects lacked the ability or capacity to resist the experimenter's suggestion. We might put this point by saying that situationism addresses situational patterns in performance, not issues in competence. (140)

Indeed, we find Brink's defense of reasons-responsiveness compelling, and to be about as good as any defense of freedom and responsibility is going to be in the face of the apparent skeptical threat posed by the situationist literature.

In arguing for his position, Brink points out that the research seems not to support the idea that in response to a range of other contexts with similar reasons, an agent would not act differently by, say, reacting to good reasons. That is to say, there is reason to think that pertinent counterfactuals about agents in the studies would support the contention that they were moderately reasons-responsive, and so did act freely and were morally responsible. Of course, the best evidence for how an agent would act in a context similar to the one she did act in is to see how she does react in such contexts. And the situationist studies were not conducted—and probably could not effectively be conducted—by exposing the same agents to variations to see how they would actually react. But as Brink reports, other agents who seemed to be relevantly like the ones in the original studies *did* react differently in pertinent variations to situation. For instance, in a situation in which the Milgram experiment was conducted and subjects were required to place the learner's hand on an electric plate to receive the "shock," only 30% administered the maximum shock as compared with 62.5% in the original experiment (Milgram 1969: 34-5, as cited in Brink 2013: 141).

We pause at this juncture to register an objection to Brink's diagnosis.²⁵ Brink contends that the "shock plate" version of the experiment suggests that subjects in the original study retained the relevant capacities. But consider the critic who contends that in the original Milgram experiment situation *does* impair capacity more in the fashion of a competence error. She might counter that the shock plate experiment elicited a different response just because the situational pressures were different, and this evidence that, in the original situation, it was extremely difficult to respond appropriately, perhaps so much so that it should be seen as exonerating (and so as undermining competence). Here, we simply wish to acknowledge the appeal of this point and proceed with our overall proposal, building on the work of Brink and others (Vargas too). Subsequent discussion will help show how we might reply.

²⁵ We are indebted to an anonymous referee for raising this concern, however, we would note that he or she does not take this to discredit Brink's point here but only to illustrate the limits of it and so to show that it is contestable. We concur, as will be illustrated in a thought experiment we will set out below, which involves two Artificial Intelligence engineers who come to different conclusions about how to conceptualize certain limits to the machine they designed. Their difference mirrors the difference here between Brink and those who would resist him in the manner explained in this paragraph.

Despite the preceding reservations, we find Brink's point is eminently plausible. Consider a number of the other studies as well. Would the seminarians in Good Samaritan not have stopped to help a person with an axe in their skull? What about one who was set ablaze or instead holding a child bleeding from the mouth? Likewise, what about the Zimbardo prison experiments? Would the "guards" have treated the "prisoners" so badly if prisoners were given the same first names as the guards' parents, or if the suggested forms of harsh treatment were degrading in even more extreme ways? Likewise for the prisoners. Would they have behaved in equally servile ways if the guards were given names like 'Hitler' or 'Goebbels', or if they were asked to pretend that the other prisoners were friends from their youth? Thinking through all of the studies, it appears that we have very little evidence that situational factors eliminate competence, and pretty good reason to think that competence is actually retained.²⁶ At the very least, we would need considerably more empirical work to draw conclusions that are any more dramatic than what is involved in noting that situational factors can play a nontrivial causal role in how a capacity is exercised, and so when it is exercised in a way that involves a performance error. Of course, it is also worth pointing out that this cuts both ways. In the Isen and Reeve experiment, what we apparently learn is that in boosting mood, situation can aid in performance *success* in the exercising of a pertinent capacity.

Manuel Vargas (2013b) offers a similar defense of a reasons-responsive view, but by slightly different means. According to Vargas, the situationist studies give reason to think that what often actually drives behavior is not a response to good reasons. It is, for instance, a response to elevation in mood, as the Isen and Reeve study suggests, or unreflective obedience to authority, as Milgram's studies suggest. Moreover, there are related worries about the pervasiveness of automaticity in general, whereby much behavior appears to be driven by nonconscious processes.²⁷ As Vargas explains:

Part of what makes automatic processes notable is not the mere fact of quick, usually sub- or unconscious mental operations but the pervasiveness of it. That is, proponents of the automaticity research program suggest that automatic behaviors are not the exception but rather the rule in human action production.²⁸

The situationist and automaticity research programs are complementary. Both emphasize that we overestimate the degree to which we understand the sources of our behavior, that conscious deliberative reflection is oftentimes affected (one might even say "contaminated") by forces largely invisible to us, and that these forces are ones that we would regard as irrelevant to the rationality of the act if we were aware of them. (Vargas, 2013b: 331)

²⁶ Perhaps there is reason for thinking that we have some degree of impairment that could *mitigate* competence somewhat. We are grateful to an anonymous referee for suggesting this qualification. The referee also points out that there is reason to be suspicious of a sharp boundary between performance failures and impaired competence. We agree, and discuss this briefly later in the paper.

²⁷ Vargas (2013b: 331) cites Kihlstrom (2008: 156), who characterizes automaticity in terms of behavior issuing from inevitable invocation of a process incorrigibly leading to completion by way of parallel processing that is screened off from other action-generating processes.

²⁸ Here Vargas sites Bargh and Ferguson (2000).

In all of these cases, Vargas argues, there is a worry that our reasons-responsive agency is bypassed (328-31). But much like Brink, Vargas then challenges the contention that such causal sources of agency show that an agent is operating in ways incompatible with retaining her capacities as a reasons-responsive agent. Sometimes, an automatic process can, for instance, be a reliable way to act in accord with morality. When it is not, it is at the very least an open question whether its role in generating behavior results merely in a performance error rather than subverting a pertinent freedom-conferring capacity (Vargas 2013b: 332).

At this juncture, we offer the further point meant to help support Brink's and Vargas's defense. They both in different ways seek to show how reasons-responsive capacities might be retained in a range of studies. In doing so, they allow us to preserve our sense that the agents in them are not affected by situational factors in a manner that degrades freedom and responsibility. A further complementary point is that in some cases, even if there *is* a reason to think that a pertinent reasons-responsive capacity is degraded, there is little reason to think that it really can be easily generalized in a way that would give rise to more pervasive skeptical worries.

Consider, for instance, the Darley and Batson Good Samaritan study. Let us suppose that a sense of urgency does degrade the seminarians' capacity for reasons-responsiveness as it pertains to helping behavior. A further question has to do with what more specific capacity is corrupted. If the capacity is merely one of failing to recognize salient natural facts in one's environment, then it will affect an agent's receptivity to reasons; agents will tend to recognize less of their surroundings and so will find few occasions for reacting morally.²⁹ A subject might fail to recognize a person slumped in a doorway, seemingly in need of medical attention, but likewise, the same subject might fail to recognize a swarm of hundred dollar bills falling from a tree, or a three legged pink giraffe. The point is that there are perfectly innocent un-generalizable effects on our ability to recognize our surroundings, and while this *can* excuse a person by way of nonculpable ignorance, the excusing here is just a normal part of life. People are often excused for innocently not seeing something which, if they saw it, would give them reason to help another in need.

Of course, a *different* explanation is that the seminarians (or some of them) did recognize the salient natural facts—they received them and internalized them as reasons—but their sense of urgency resulted in their not reacting to such facts. It is a further question, however, whether in these cases we have a performance error, in which case the agents still seem to be culpable, or instead a case of compromised capacity that would be grounds for exculpation. In the latter case, there seems to be a greater cause for concern. If the capacity to *react* to moral reasons for helping can be so easily degraded by situations involving urgency, this would be more troubling and more suggestive of a skeptical threat to our regarding others in a range of relevant contexts as free and responsible.

In closing this section, we endorse the basic defense of reasons-responsiveness advanced by Brink and Vargas. Nevertheless, we also note an important caveat to which it appears both Brink and Vargas are sensitive, as is Nelkin. Whether the proposed reasons-responsive theorists' diagnosis is correct depends on empirical details that are left unanswered. Brink is correct that the situationist

²⁹ Nelkin makes a similar remark about what might explain the failure to help in the Good Samaritan experiment.

studies on offer provide little reason to conclude that we have breakdowns in relevant capacities. For the most part, all we have from the studies themselves is evidence of performance failures. But it might be that the hidden causal details will after all expose episodic capacity failures.³⁰ As Vargas observes (2013b: 330), we will only be able to tolerate so much irrationality in performance before we begin to worry that the better explanation unifying the range of situationist (and automaticity) studies is that the capacities at issue are after all degraded by situation. Nevertheless, the evidence for this more dramatic worry is thin, and it also flies in the face of such a wide range of conduct in everyday life that can be explained by the presupposition that most healthy, sane adults are by and large moderately reasons-responsive.

6. Pessimistic Worries that Remain

We now wish to qualify our conclusions drawn in the last section. So as to avoid any misimpression, we are not retracting our contention that, given the present state of the situationist literature, the reasons-responsive theorists have an adequate way to defend against the apparent skeptical threat to freedom and responsibility. That is, we are committed to free will and moral responsibility realism, and nothing in what follows is intended as a retreat from that commitment. We state this explicitly here since the point we wish to make now can easily be misunderstood as endorsing a strong skeptical conclusion.³¹

In order to tease out our remaining pessimistic worries, imagine two AI engineers, Geno and Georgiana. Geno and Georgianna work for an auto repair company, and their job is to develop a new and improved machine sensitive to emissions problems in cars. Drive a car over a ramp as you move through any gas station, and the new gizmo will diagnose a vast array of emissions problems with the car, rule out false positives if the car has bad gas but otherwise functions well, pinpoint what sorts of emissions problems are at issue, and suggests repairs and their urgency. The machine is, in a sense, a limited reasons-responsive “agent.” Indeed, we can even attribute to it receptivity and reactivity characteristics. Maybe sometimes, when another car is too close, it gets “noise” that looks like a good reason for a diagnosis but isn’t. And suppose the AI system is reasonably able to sort this out and eliminate the bad noise. So it has “receptive” resources. Also, suppose it is able to yield different reports by telling the operator of the car that the problem is not too bad and the car would still pass an emissions test but is only less than optimally efficient. Or instead it can report that the car has a vapor leak and is dangerous to drive, and so on. Thus it can be differentially reactive. Suppose also that it is able to “learn,” maybe getting signals sent from computer chips in the cars so that it can upgrade and so on.

Now suppose that as Geno and Georgianna are designing the AI system, they discover various foibles with its operation. It won’t work so well when, say, the winds are over 20 mph, since the winds move car emissions too quickly for reliable detection. Or instead it has a hard time

³⁰A further wrinkle, one we’ll not develop here, is that some situational factors, rather than undermining or degrading a capacity might instead give rise to or bolster it. That is, there is an assumption here, in Nelkin’s formulation of the apparent threat, that situational factors are problematic because they are freedom-defeating. This might be true, but of equal interest is the prospect that other situational factors are freedom-sustaining. We are indebted to Hannah Tierney for calling this to our attention.

³¹ For a view that is similar in its findings, see Nahmias (2007).

discerning a poorly functioning 1999 Ford F150 from an excellent functioning 2002 Chevy S10. Now imagine that Geno can get the AI to do better with the F150s but only by doing worse with the Chevy S10, and Georgianna can get the opposite results. She can tweak the system to do better with the Chevy but worse with the Ford. Here, Geno and Georgianna might be seen as arriving at equally attractive options that, while less than optimal, are nevertheless “patches” that are helpful and are the best fixes that could be expected given the current state of technology. While Geno and Georgianna have various options (do nothing, improve the response to the Ford, or improve the response of the Chevy), and while the options are less than they might have hoped for, they can plausibly regard their results as satisfying given expectations about the current state of technology. Of course it is imaginable that they could have built a more perfect AI system, but what they have by reasonable standards is satisfying.

The analogy here to theories of free will might be this: In developing a theory of free will and moral responsibility, the theory of freedom on offer, if it is to treat in realist fashion creatures like us here on earth, will have to allow for some foibles or limitations. We human beings act recklessly. We sometimes act from weakness of will. Some of us are more or less intelligent than others, more or less prudent, have greater resources for perseverance, and so on. If we were to imagine “building” a critter like one of us with the aim of her being able to act freely, we would not want to set our standards for the requirements on agency too high. Otherwise, most of us who operate in ways that fall shy of an ideal kind of agency would not be regarded as free and responsible in a considerable sphere of our lives. Our aspirations as theorists about free will, therefore, should, like Geno’s and Georgianna’s, not be too exaggerated, lest they unnecessarily give rise to skeptical or pessimistic conclusions. Perfect agency (whatever that is) should not be made the enemy of good-enough agency. Thus, our theories of free will and moral responsibility should not make *a priori* demands for an extremely high-grade kind of agency, one that would require a kind of reasons-responsiveness that is never (or only very rarely) susceptible to performance errors due to situational features.

But now, suppose that Geno and Georgianna discover that there are some really weird glitches in the system. Suppose they discover that if two children are in the back seat of any car reciting nursery rhymes, the AI system will yield bad results. Or instead, suppose they learn that if the car being tested is red and there are uncut blades of grass within twenty feet of the AI’s sensors, they will get bad results.³² Imagine that Geno and Georgianna even learn the unexpected causes whereby the system gets these bad results, but, as it happens, they are largely uncorrectable. (Of course, they could require that no children recite nursery rhymes while the system is in operation, and so on, just like we could try to have all people avoid ever being told what to do by people in positions of authority.) In such a case, suppose that Geno concludes that the AI system temporarily loses a capacity, whereas Georgianna concludes that the system in those cases spits out a performance error that is caused by the bizarre situational factor. Georgianna might reason with Geno this way:

³² We can also stipulate that there are other kinds of weird glitches, ones that actually *improve* the AI’s sensing capabilities. For example, if the driver has sneezed in the car that day, the AI sensors will spit out *extremely accurate* information about the tread and pressure of the back tires. The sensors were not designed to do this, but as it turns out these obscure situational features actually trigger increased diagnostic capabilities.

Look Geno, given the AI tools we have, we knew that we'd never get perfect reliability. We knew that we'd get failures at points. But we designed this AI system to operate and respond to various bits of information in its environment, and as it happens, the subsystems as we have devised them are after all in play. What we have is a system that is operating and making mistakes, but these are mistakes that do not involve the *incapacitation* of the system. It is just making an error. If we want something to do this work, this is the best the world has to offer us right now, damn it all!

We hope the point of our analogy is clear. Whereas Geno has positioned himself much like a situationist skeptic about this analogue to free will and moral responsibility, Georgianna is, basically, in a position analogous to the reasons-responsive *realist* responding to the alleged threat of the situationist literature. And Georgianna's reply to Geno is much like the reply we have proposed for reasons-responsive theorists.

While we think that Georgianna's reply to Geno is appropriate, we also think that it would not be unreasonable for Georgianna to adopt a pessimistic—but *not skeptical*—view about the quality of the AI system, especially given that it uses resources that are about the best that we can hope for. The realist reply to the situationist literature that we have so far been pushing is that we human agents, if we are after all free and responsible, and if situational features do not pose a great threat to our *capabilities*, are just plain built a lot more poorly than we expected. Supposing the experimental results from the situationist literature are accurately reported, the causal sources of our performance errors (and successes) reveal disturbing limitations on what can be expected of us as morally responsible agents. This, some might think, is a rational basis for a kind of pessimism about the quality of human agency.

While this might not undermine our freedom and responsibility as such, it casts what freedom and responsibility we have in a relatively unflattering light. This is the source of our pessimistic worry: even if our reasons-responsive capabilities are not impugned by the situationist literature to the point where they undermine freedom and responsibility, our picture of our own agency is considerably diminished if, for example, as it happens our moral concern for others can so easily be shifted about depending on utterly trivial alterations in conditions, say, involving colors or smells in the room.³³

Whereas, given our folk conception of ourselves as free agents, we have a picture of our rational selves suggesting that we humans at our best are like the finest well-tuned twenty-first century performance cars, the empirical findings of the situationists suggest that we are after all built a lot more like Model Ts. This is disheartening and a cause for an empirically informed and limited kind of pessimism about the human condition. It is not that *uniformly* we perform a lot less reliably due to these sorts of odd situational factors. It is, rather, that often in various patchwork ways we are far less reliable than we would like to be, and we would have thought that as reasons-responsive beings we were built a lot better than apparently we are.

³³ Miller (2013: ch. 2) builds on similar considerations to argue for the claim that most people do not possess traditional virtues such as compassion. That most of us can be “pushed around” by situational features in this way suggests that we have what he calls “mixed traits,” traits that are neither entirely good nor bad in any given situation.

What should we make of this pessimistic worry? Here we grant that reasonable minds can differ. Some might feel that, in opposition to our proposal, this sort of pessimism *is* grounds for skepticism. Freedom so built, if this is really what it comes to for us, is just not freedom enough. This would be, so to speak, to side with Geno rather than Georgiana and insist that once we see what sorts of causes are in play here, what we should think is that there is some sort of episodic lack of competence. Derk Pereboom (2001, 2014) argues for skepticism about a certain kind of freedom and moral responsibility (desert entailing) in Spinozistic fashion by seeking to expose the hidden causes of our putatively free acts. In doing so, he means to elicit a judgment that agents so caused lack a capacity for the sort of freedom that matters for responsibility. Pereboom's arguments are metaphysical and meant to be fully general: Either determinism is true, and tending to the hidden causes of our behavior shows that we are all unfree in the relevant way, or indeterminism is true, and tending to the hidden causes of our behavior shows that we are all unfree in the relevant way. But Pereboom's key strategy could be used in this more limited context as well. If we come to see that the hidden causes involved in our reasons-responsive "performance errors" are frequently the product of situational factors, much of the freedom that we allegedly find is shown to be benighted and so not adequate grounds for deserving praise or blame. Why? Because what is *really* exposed is some sort of failure of competence. Our freedom in this way hangs too much on situational luck, and so, in the spirit of Pereboom's Spinozistic strategy, the pessimism we have identified here should give way to skepticism when relevant situational factors are in play.³⁴

Yet another intriguing response is the sort that Vargas (2013b) advocates, requiring a revisionary picture of what agency is, and so what free and responsible agency are. Agents can act freely and be morally responsible, but grounds for a non-skeptical treatment of freedom in the face of the situationist literature require that we understand the conditions of agency as varying in different

³⁴ There is one other possibility we will not develop here but is worth considering. One might press for skepticism by instead accepting that the pertinent situational factors affect performance errors *without* undermining a capacity. On this skeptical strategy, the skeptic resists the realist by pointing out that the hidden causes involved in the exercise of the capacities when there are performance errors shows that the mere possession of and use of these capacities is inadequate for freedom and responsibility. This would, in effect, be like arguing that there is a further category of excuse beyond the traditional ones friendly to a compatibilist proposal like Brink's. What defeats freedom and responsibility, given this category, is not any failure of capacity but instead a set of underlying causes that expose the inadequacy of the sources of the exercising of these capacities. In this case, the pessimism we propose would provide good reason for something further: skepticism.

There is an interesting related question here about how to understand Pereboom's (2001, 2014) metaphysical argument against compatibilism. In his famous Four Case Argument for incompatibilism, he structures a case of manipulation in which all the (putative) compatibilist sufficient conditions are in place. He then uses a manipulative source giving rise to these conditions to show that the underlying causes undermine our judgment of freedom and responsibility. One way to interpret his argument is that the manipulation reveals that some freedom *capacity* is corrupted, and so a conventional excusing consideration is after all in play. This is basically the strategy we considered above as a way for the skeptic to argue. But another way to interpret Pereboom's argument is that the manipulation reveals that no freedom *capacity* at all is corrupted, and so no traditional compatibilist-friendly excuse applies; rather, the causes involved in *exercising* the pertinent capacities are shown not to provide a satisfying basis for freedom and responsibility. If so, the excusing consideration here is of a different sort—a further category—beyond those that work by showing that a capacity of one sort or another (say, epistemic or control) is defeated.

We are indebted to Dana Nelkin for extremely insightful comments regarding these issues.

circumstances. There are, so to speak, agents-in-conditions-responding-to-authority and agents-when-there-is mood-boost, and so on. But there just are not agents as stable beings whose nature and capacities for free agency are invariant across variation in circumstances. That is what we should be skeptical about, and the situationist literature provides a basis for this skepticism. Fortunately, Vargas argues, there is a better way to reconceive our agency, and in doing so, we can resist full-blown skepticism.

We propose a different response here. Our contention, insofar as a non-skeptical and non-revisionary response is called for, is that the situationist literature simply shows our human agency to be limited albeit consistent with realism, and like Georgianna's estimation of her own AI system, there is a reason to regret that our agency can't be better than it is and provide us with a more effective kind of freedom. In this respect, our proposal is relevantly like Paul Russell's (2013, provisionally forthcoming 2016) brand of pessimistic compatibilism. On his view, if we attend to the ultimate sources of our agency, we will recognize a kind of limit or luck in how it is that we are free and how it is that we are responsible. It is not that these limits show us not really to be free and responsible. But it does "put the lie" on our place in the natural world, and it reveals the freedom we do have as limited or hamstrung in a way that is disappointing given what we might have hoped for. Now Russell offers this proposal in response to the general metaphysical problem about free will in the face of determinism or instead indeterminism (just as Pereboom offers his proposal). Ours is, again, not meant to apply as such. It is, rather, a response to empirical results that suggest that these worries, while not universal but instead piecemeal, are widespread, so that often our agency is a lot less effective than we would have hoped it to be.

Recall that early on we noted that the prospects for a reasons-responsive theory, as well as the skeptical threats from the situationist literature, are not limited to compatibilism. Incompatibilists requiring libertarian satisfaction conditions for freedom and responsibility are open to the same theoretical prospects and skeptical threats. We mean for this to apply to the sort of pessimism we have identified here. Call ours an empirically grounded, piecemeal *pessimistic realism*—whether compatibilist or libertarian—about freedom and responsibility in response to the situationist literature.

We wish to point out that the viability of our pessimistic realism strategy for responding to the situationist threat to freedom and responsibility is sensitive to certain empirical facts. That is, we concede that there are things that we could discover about human agency that would defeat our strategy. What kinds of things? Well first, suppose we discovered that small situational features were pretty much *always* adversely affecting our MRR capacities. That is, our best empirical literature might reveal that even if we retained our reasons-responsiveness competence, if situational features were ubiquitously causing performance errors, then this would be problematic for our strategy. If all that the MRR theorist could offer in the face of such evidence is to say, "Don't worry, the competence remains!", we think that this would be cold comfort indeed. Even more obviously problematic for our pessimistic realism strategy would be the discovery that situations were ubiquitously causing *competence* errors. Our contention (along with Brink, and we think in the spirit of Nelkin and Vargas) has been that the empirical literature at best shows that situational features sometimes causally contribute to performance errors in responding to reasons, but that the relevant competence remains. It is this remaining competence that can ground freedom and responsibility, even when performance errors are made. But if this

competence was severely and ubiquitously incapacitated, this would knock out the legs from our strategy.

Of course, much of this previous discussion relies on having a fairly clear notion of what counts as a performance error and what counts as a competence error when it comes to MRR. It would be crucial to know, for example, at what point the presence of a lot of performance errors due to situational features adds up to (or provides good evidence of) a failure of competence in responding to reasons. Furthermore, the challenges to our pessimistic realism just adumbrated rely on the empirical literature turning out a certain way: revealing either that situational features result in ubiquitous performance or competence errors. But not only does the current research not suggest this picture, we doubt that such a picture will be vindicated. After all, people like you and us get around in the world quite successfully every day: operating cars, doing creative and meaningful work, buying carrots to make stew, raising well-adjusted children, sending people into space, voting our consciences, and so on. The manifest image of our place in the world strongly suggests that most of us *do* tend to be moderately reasons-responsive, and to be so in ways that are largely successful and desirable. We stress again that this picture is answerable to the empirical facts on the ground. But as things stand, we remain unconvinced that situational features play the kind of competence-undermining role that would require us to give up on the view that we are free and morally responsible because situational features undermine the relevant capacities to respond to reasons.

7. Why Character Matters (at Least Indirectly)

Throughout, we have focused on the relationship between the situationist literature and free will and moral responsibility. We offered a progress report: At present the empirical literature does not give us good enough reason to think that we lack freedom and moral responsibility on the grounds that situational features undermine our reasons-responsive capabilities. This is our realism. However, we also suggested that the situationist literature should give us pause. How so? Our reasons-responsive capabilities may be a lot more poorly put together than we had expected or hoped. Even if most of us usually meet the threshold requirements for having free will and moral responsibility, we still might fall much closer to the minimum threshold than we'd like. This is our pessimism.

We conclude, however, on a more optimistic note. Thus far, we have said very little about the relationship between situationism and character, which has traditionally been thought of as the primary target of the situationist research program. Whatever the outcome of that debate, consider the following possibility. Suppose that our somewhat pessimistic assessment of our agency is, in light of the situationist literature, largely accurate: most of us don't typically fall below the threshold for morally responsible agency, but we are also not typically very high above it. Our reasons-responsive capabilities are, more often than not, perhaps closer to being a Model T than a Porsche 911. But *must* we retain a kind of Model T agency? Can we do better?

Here, we briefly suggest an optimistic plot twist: that we might diminish this aforementioned pessimism by showing that we can cultivate character traits that prevent situational features from inhibiting successful deployment of MRR. The story might go like this. Character traits are kinds of dispositions. And as we pointed out in Section 1, we are complex psychological beings:

different traits can work against each other such that one suppresses another. Suppose that in the absence of the possession of some trait *T*, when we see people slumped in sidewalks while in a hurry, we typically fail to stop and offer aid. This might be due, let us suppose, to certain mixes of situational factors gumming up our MRR capabilities and issuing in a kind of performance error. But now, suppose that if trait *T* was *also* in the mix, *T* would typically serve to prevent that gumming up of the works. So understood, the possession of *T* would typically prevent the performance error. This is, then, a way of seeing how character traits might be helpful in combatting the corrosive effects that situational features might have on our agency.

Of course, in order to improve reasons-responsiveness overall or in limited domains, character traits need not always prevent such errors, and even when they do, they need not do so in the same ways. One way character traits might do so is that, once in a situation where certain features threaten to degrade our agency, a trait could prevent those features from having those effects. Another way that character traits might combat the corrosive effects of situational features is that they could help us avoid altogether being in situations that can degrade our freedom (either by compromising our capacities or instead playing a causal role in our performance errors).³⁵ Alternatively, character traits could operate in a corrective manner—even when we do find ourselves in situations that threaten to thwart desired moral outcomes—by outweighing or counter-balancing the effects that those situational features have on our psychologies.³⁶ Furthermore, the way character traits may boost our MRR need not be all or nothing. Suppose that, absent trait *T*, most of us would fail to help the slumping person when in a hurry due to a performance error at about a 60% rate. Yet with *T*, that failure rate would drop to 40%. This would amount to a considerable improvement in our agency even if trait *T* would not prevent all performance errors in the relevant situations.

This is all well and good, but is there any reason to think that character traits can work this way? Perhaps so, especially if research from the psychological sciences could tell us what kinds of traits—understood as dispositions—could help us. Something to this effect has been recently suggested by Alfred Mele and Joshua Shepherd (2013) who argue that “knowledge about the impact of situations on behavior can actually boost agents’ power to counteract harmful situational effects” (62). Here is one way to think about the boost, using an example they cite from the psychological literature on the efficacy of so-called implementation intentions.³⁷

To form an implementation intention involves a person committing himself or herself to respond to a certain situation in a certain manner (Gollwitzer 1999: 494). Such intentions are subordinate to goal intentions and specify *how*, *when*, and *where* the goal is met. One study of the effect of implementation intentions involves drug addicts who showed symptoms of withdrawal (496). One group of addicts was asked in the morning to form the goal intention of writing a short

³⁵ Doris (2002) himself argued that by becoming good at choosing situations we can counteract the corrosive effects of situational features on our desired moral outcomes. Doris thought such a skill was compatible with his own skepticism about global traits, but see Rodgers and Warmke (2015) for an alternative view.

³⁶ We thank Christian Miller for this suggestion.

³⁷ Here, we will focus just on the ways that implementation intentions may help boost MRR. There may of course be other strategies for strengthening traits in order to resist the negative influences of situational features. See, for example, Miller (forthcoming).

curriculum vitae before 5:00 p.m. They were also tasked with forming implementation intentions about when and how they would complete the task. The second group was asked to form the same goal intention but was not told to plan how to accomplish that goal. Mele and Shepherd report the striking results: “although none of the people in the second group completed the task, 80% of the people in the first group completed it” (2013: 77).³⁸

What the efficacy of implementation intentions suggests is that dispositions formed at an earlier time can be efficacious later on down the road when we are faced with certain kinds of situational influences or pressures. And if we think of character traits as dispositions or constellations of dispositions, then we can see why there is reason to hope that the taking on of certain kinds of traits can help us to avoid the negative influences that situational features may pose to our agency. Our reasons-responsiveness capacities may never be as good as we hoped. But we think that, not only are they not as bad as some skeptics have warned, but there is good reason for a limited optimism that such capacities can be improved.

So, we close with a glimmer of optimism, even in light of our pessimistic realism. Many would want more, but in our estimation, they are holding us up to expectations for a high-grade form of agency that the science of psychology suggests we lack. Whether somehow creatures like us *could* live up to such expectations is another question altogether. All the sciences have shown us so far is that we typically do not. Of course, as an aspirational goal, it might be best to act as if we are able to do so even if as it turns out we are simply incapable of doing so.

Bibliography

Asch, Solomon. 1951. “Effects of Group Pressures upon the Modification and Distortion of Judgment.” In *Groups, Leadership, Men*. Ed., Harold Guetzkow. Pittsburgh: Carnegie Press: 177-90.

Bargh, John A., and M.J. Ferguson. 2000. Beyond Behaviorism: On the Automaticity of Higher Mental Processes. *Psychological Bulletin* 126 (6 Special Issue): 925-45.

Blass, T. 1996. “Attribution of Responsibility and Trust in the Milgram Obedience Experiment,” *Journal of Applied Social Psychology* 26:1529–1535.

Bowers, K. 1973. “Situationism in Psychology: An Analysis and a Critique,” *Psychological Review* 80: 307-36.

Brink, David. 2013. “Situationism, Responsibility, and Fair Opportunity,” *Social Philosophy and Policy* 30: 121-149.

Brink, David and Nelkin, Dana. 2013. “Fairness and the Architecture of Responsibility.” In David Shoemaker, ed., *Oxford Studies in Agency and Responsibility*. Vol. 1 Oxford: Oxford University Press: 284-313.

³⁸ Mele and Shepherd note that many such studies are discussed in Gollwitzer (1999) and that Gollwitzer and Sheehan (2006: 69) report that “findings from 94 independent tests showed that implementation intentions had a positive effect of medium-to-large magnitude...on goal attainment.”

- Burton, R. 1963. "Generality of Honesty Reconsidered," *Psychological Review* 70:481–499.
- Ciurria, Michelle. 2013. "Situationism, Moral Responsibility and Blame," *Philosophia* 41: 179-193.
- Darley, J. 1995. "Constructive and Destructive Disobedience: A Taxonomy of Principal-Agent Relationships," *Journal of Social Issues* 51:125–154.
- Darley, J. and C. Batson. 1973. "'From Jerusalem to Jericho:' A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27: 100-108.
- Dennett, Daniel. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
- Doris, John. 2002. *Lack of Character*. Cambridge, Mass.: Cambridge University Press.
- Doris, J., Murphy, D. 2007. "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity," *Midwest Studies in Philosophy* 31:25–55.
- Elms, A. 1995. "Obedience in Retrospect," *Journal of Social Issues* 51:21–31.
- Epstein, S. 1979. "The Stability of Behavior: I. On Predicting Most of the People Much of the Time," *Journal of Personality and Social Psychology* 37: 1097-126.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.
- Flanagan, Owen. 1991. *Varieties of Moral Personality*. Cambridge: Harvard University Press.
- Gert, Bernard, and Tim Duggan. 1979. "Free Will as the Ability to Will," *Nous* 13: 197-217.
- Goldhagen, D. 1996. *Hitler's Willing Executioners: Ordinary Germans and the Holocaust*. Knopf, New York.
- Gollwitzer, P. 1999. "Implementation Intentions," *American Psychologist* 54: 307-316.
- Gollwitzer, P. and P. Sheeran. 2006. "Implementation Intentions and Global Achievement: A Meta-Analysis of Effects and Processes," *Advances in Experimental Social Psychology* 38: 69-119.
- Haji, Ishtiyaque. 1998. *Moral Appraisability*. New York: Oxford University Press.
- Haney, Craig, Curtis Banks, and Philip Zimbardo. 1973. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology and Penology*. 1: 69-97.
- Haney C, Zimbardo P. 1998. "The Past and Future of U.S. Prison Policy: Twenty-five Years after the Stanford Prison Experiment," *American Psychologist* 53:709–727.

- Harman, Gilbert. 1999. "Moral Psychology Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99: 315-31.
- Isen, A. and P. Levin. 1972. "Effect of Feeling Good on Helping: Cookies and Kindness," *Journal of Personality and Social Psychology* 21: 384-388.
- Kamtekar, Rachana. 2004. Situationism and Virtue Ethics on the Content of Our Character. *Ethics* 114: 458-91.
- Kihlstrom, John F. 2008. *The Automaticity Juggernaut—Or, Are We Automotons After All?* In *Are We Free? Psychology and Free Will*, edited by John Baer, James C. Kaufman, and Roy F. Baumeister. New York: Oxford University Press.
- Latane B, Darley J. 1970. *Unresponsive Behavior – Why doesn't he help?* New York: Appleton-Century-Crofts.
- Latané, B. and J. Rodin. 1969. "A Lady in Distress: Inhibiting Effects of Friends and Strangers on Bystander Intervention," *Journal of Experimental Social Psychology* 5: 189-202.
- Lutsky, N. 1995. *When is 'obedience' Obedience? Conceptual and Historical Commentary*, *Journal of Social Issues* 51:55–65.
- McKenna, Michael. forthcoming. "Reasons-Responsive Theories of Freedom." *Routledge Handbook on Free Will*. Routledge Press.
- _____. 2013. "Reasons-Responsiveness, Agents, and Mechanisms." In David Shoemaker, ed., 2013. *Oxford Studies in Agency and Responsibility*, vol. 1. Oxford: Oxford University Press: 151-84.
- Mele, Alfred. 1995. *Autonomous Agents*. New York: Oxford University Press.
- Mele, Alfred and Joshua Shepherd. 2013. "Situation and Agency." *Journal of Practical Ethics* 1: 62-83.
- Milgram, S. 1969. *Obedience to Authority*, New York: Harper & Row.
- Miller, A. 1995. "Constructions of the Obedience Experiments: A Focus Upon Domains of Relevance," *Journal of Social Issues* 51:33–53.
- Miller, Christian B. 2013. *Moral Character: An Empirical Theory*, Oxford: Oxford University Press.
- Miller, Christian B. 2014. *Character and Moral Psychology*, Oxford: Oxford University Press.
- Miller, Christian B. 2016. "Situationism and Free Will," *Rutledge Companion to Free Will*, Ed. Meghan Griffith, Kevin Timpe, and Neil Levy. Routledge.

- Miller, Christian B. forthcoming. "Virtue Cultivation in Light of Situationism," in Julia Annas, Nancy Snow, and Darcia Narvaez, *Developing the Virtues: Integrating Perspectives*. Oxford: Oxford University Press.
- Modigliani, A., Rochat, F. 1995. "The Role of Interaction Sequences and the Timing of Resistance in Shaping Obedience and Defiance to Authority.," *Journal of Social Issues* 51:107–123.
- Nahmias, Eddy. 2007. "Autonomous Agency and Social Psychology." In Maraffa, De Caro, and Ferriti, Eds., *Cartographies of the Mind*, Vol. 4, Dordrecht: Springer: 169-85.
- Nelkin, Dana. 2011. *Making Sense of Freedom and Responsibility*. Oxford University Press.
- _____. 2005. "Freedom, Responsibility, and the Challenge of Situationism." *Midwest Studies in Philosophy* 29: 181-206.
- Newcomb, T. 1929. *Consistency of Certain Extrovert-introvert Behavior Patterns in 51 Problem Boys*. Columbia University, Teachers College, Bureau of Publications, New York.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- _____. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Rochat, F., Modigliani, A. 1995. "The Ordinary Quality of Resistance: From Milgram's Laboratory to the Village of Le Chambon," *Journal of Social Issues* 51:195–210.
- Rodgers, Travis J. and Brandon Warmke. 2015. "Situationism versus Situationism," *Ethical Theory and Moral Practice* 18: 9-26.
- Rosenthal, A. 1999. *Thirty-eight Witnesses: The Kitty Genevese Case*, 2nd edn. Berkeley: University of California Press.
- Ross, L. and R. Nisbett. 1991 *The Person and the Situation: Perspectives of Social Psychology*, New York: McGraw-Hill.
- Russell, Daniel. 2009. *Practical Intelligence and the Virtues*, Oxford: Clarendon Press.
- Russell, Paul. provisionally forthcoming. "Free Will Pessimism." In David Shoemaker, ed., 2016. *Oxford Studies in Agency and Responsibility*, vol. 4, Oxford: Oxford University Press.
- _____. 2013. "Responsibility, Naturalism, and the Morality System." In David Shoemaker, ed., 2013. *Oxford Studies in Agency and Responsibility*, vol. 1, Oxford: Oxford University Press: 184-204.
- Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press.
- Sears, R. 1963. "Dependency Motivation." *Nebraska Symposium on Motivation*, 11: 25-64. University of Nebraska Press: Lincoln.

- Sreenivasan, G. 2002. "Errors about Errors: Virtue Theory and Trait Attribution," *Mind* 111:47–68.
- Talbert, Matthew. 2012. "Moral Competence, Moral Blame, and Protest." *Journal of Ethics* 16: 89-109.
- _____. 2009. "Situationism, Normative Competence, and Responsibility for Wartime Behavior." *Journal of Value Inquiry* 43: 415-32.
- Talbert, Matthew and Jessica Wolfendale. 2015. "War Crimes: Causes, Excuses, and Blame," ed. Christian Miller, et al., *Character: New Directions from Philosophy, Psychology, and Theology*. New York: Oxford University Press.
- Van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- Vargas, Manuel. 2013a. *Building Better Beings*. Oxford, UK: Oxford University Press.
- _____. 2013b. "Situationism and Moral Responsibility: Free Will in Fragments," in Andy Clark, Julian Kiverstein, and Tillmann Vierkant (eds.), *Decomposing the Will*, Oxford: Oxford University Press, pp. 325-349.
- Watson, Gary. 1987. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In Ferdinand Schoeman, ed. 1987. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press: 256-86.
- Wolf, Susan. 1990. *Freedom within Reason*. Oxford: Oxford University Press.
- _____. 1987. "Sanity and the Metaphysics of Responsibility." In Ferdinand Schoeman, ed. 1987. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press: 46-62.
- Zimbardo, P. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*. Random House Trade Paperbacks.
- Zimbardo, Philip. <http://prisonexp.org>.