Forthcoming in *Decomposing the Will*. Edited by Till Vierkant, Julian Kiverstein and Andy Clark. New York: Oxford University Press. Draft as of 08/24/11.

Situationism and Moral Responsibility: Free Will in Fragments

Manuel Vargas

Many prominent accounts of free will and moral responsibility treat as central the ability of agents to respond to reasons. Call such theories *Reasons* accounts. In what follows, I consider the tenability of Reasons accounts in light of situationist social psychology and, to a lesser extent, the automaticity literature. In the first half of this chapter, I argue that Reasons accounts are genuinely threatened by contemporary psychology. In the second half of the paper I consider whether such threats can be met, and at what cost. Ultimately, I argue that Reasons accounts can abandon some familiar assumptions, and that doing so permits us to build a more empirically plausible picture of our agency.

I. Preliminaries: free will and moral responsibility

'Free will' is a term of both ordinary and technical discourse. We might say that Ava chose to major in mathematics "of her own free will" or that Ruth lacks free will because of, say, brainwashing or mental illness. What is less clear is whether ordinary sorts of usages of free will reflect a unified or single concept of free will. Perhaps ordinary usage picks out distinct features of the world, masquerading under a single term. Referential fragmentation seems an especially appealing hypothesis when one considers that varied characterizations given to free will among scientists, philosophers, and theologians. So, for example, scientists have used the term 'free will' to refer to, among other things: the feeling of conscious control (Wegner 2002); "*undetermined choices* of action," (Bargh 2008, p. 130) and the idea that we choices "independent of anything remotely resembling a physical process" (Montague 2008, p. R584-85). Philosophical uses display variation, too. Among other things, free will has been characterized as: the ability to do otherwise; a kind of control required for moral responsibility; decision-making in accord with reason; and a capacity to act in the way we believe when we deliberate about what to do. The univocality of 'free will' is dubious (Vargas 2011).

In what follows, I treat free will as the variety of control distinctively required for agents to be morally responsible.[1] It is a further matter, one I will not address here, whether such control constitutes or is a part of any other powers that have sometimes been discussed under the banner of 'free will'. Among theories of free will characterized along these lines, of special interest here are *Reasons* accounts. These are accounts on which an agent's capacity to appropriately respond to reasons constitutes the agent's having the form of control that (perhaps with some other things[2]) constitutes free will or is required for moral responsibility (Wolf 1990; Wallace 1994; Fischer and Ravizza 1998; Arpaly 2003; Nelkin 2008). There are a number of attractions to these accounts, although here I will only gesture at some of them.[3]

First, Reasons theorists have been motivated by the idea that in calling one another to account, in (especially) blaming one another and in judging that someone is responsible, we are suggesting that the evaluated agent had a *reason* to do otherwise. Having reams of alternative possibilities available, even on the most metaphysically demanding conception of these things, is of little use or interest in and of itself. It is a condition on the possibility of an alternative being relevant that there be a reason in favor of it, and this is true for both the purposes of calling someone to account, and for an agent trying to decide what to do. In the absence of the ability to discern or act on a discerned reason in favor of that possibility, it is something of an error or confusion to blame the agent for failing to have acted on that alternative (unless, perhaps, the agent knowingly undermined or destroyed his or her reasons-responsive capacity).

Second, and relatedly, Reasons accounts appear to cohere with the bulk of ordinary judgments about cases (e.g., why young children are treated

---

1 In the contemporary philosophical literature on free will, this seems to be the dominant (but not exclusive) characterization of free will (Vargas 2011).

2 These "other" conditions might be relatively pedestrian things: for example, being sometimes capable of consciousness, having beliefs and desires, being able to form intentions, having generally reliable beliefs about the immediate effects of one's actions, and so on. Also, a Reasons account need not preclude other, more ambitious demands on free will. One might also hold that free will requires the presence of indeterminism, non-reductive causal powers, self-knowledge, etc.. However, these further conditions are not of primary interest in what follows.

3 Elsewhere, I have attempted to say a bit about the attractions of a Reasons view in contrast to "Identificationist" views (Vargas 2009). There are, however, other possibilities beyond these options.

differently than normal adults, why cognitive and affective defects seem to undermine responsibility, why manipulation that disrupts people's rational abilities seems troublesome, etc.). So, there is a "fit" with the data of ordinary practices and judgments.

Finally, Reasons accounts provide us with a comparatively straightforward account of our apparent uniqueness in having free will and being morally responsible. To the extent to which we are responsive to a special class of considerations (and to the extent to which it is worthwhile, valuable, or appropriate to be sensitive to these considerations) this form of agency stands out against the fabric of the universe; it constitutes a particularly notable form of agency worth cultivating. Reasons accounts are thus appealing because of a package of explanatory and normative considerations.

However we characterize reasons, it would be enormously problematic if we seldom acted for reasons, or if it turned out that there was a large disconnect between conscious, reasons-involved deliberative powers and the causal mechanisms that move us. Unfortunately, a body of research in social psychology and neuroscience appears to suggest exactly these things (Doris 2002; Nelkin 2005; Woolfolk et al. 2006; Nahmias 2007).

2. Situationism

Consider the following classic social psychology experiments:

> **Phone Booth:** In 1972 Isen and Levin performed an experiment on subjects using a pay phone. When a subject emerged from the pay phone, confederates of the experimenters "inadvertently " spilled a manila envelope full of papers in front of the subject as the subject left the phone booth. The remarkable thing was the difference a dime made. When subjects had just found a dime in the change return of the telephone, helping behavior jumped to almost 89% of the time. When subjects had not found a dime in the change return, helping behavior occurred only 4% of the time (Isen and Levin 1972).[4]
>
> **Samaritan:** In 1973, Darley and Batson performed an experiment on seminary students. In one case, the students were asked to prepare a talk on the Good

4 It may be worth noting that the particulars of this experiment have not been easily reproduced. Still, I include it because (1) there is an enormous body of literature that supports the basic idea of "mood effects" dramatically altering behavior and (2) because this example is a familiar and useful illustration of the basic situationist idea. Thanks to Christian Miller for drawing my attention to some of the troubles of the Isen and Levin work.

Samaritan parable. In the other case, students were asked to prepare a talk on potential occupations for seminary students. Subjects were then told to walk to another building to deliver the talk. Along the route to the other building, a confederate of the experimenters was slumped in a doorway in apparent need of medical attention. The contents of the seminarian's talks made little difference in whether they stopped to help or not. What did have a sizeable difference was how time-pressured the subjects were. In some conditions, subjects were told they had considerable time to make it to the next building, and in other conditions subjects were told they had some or considerable need to hurry. The more hurried the subjects were, the less frequently they helped (Darley and Batson 1973).

Obedience to Authority: In a series of widely-duplicated experiments, Stanley Milgram showed that on the mild insistence of an authority figure, a range of very ordinary subjects were surprisingly willing to (apparently) shock others, even to apparent death, for failing to correctly answer innocuous question prompts (Milgram 1969).

The literature is filled with plenty of other fascinating cases. For example: psychologists have found that members of a group are more likely to dismiss the evidence of their senses if subjected to patently false claims by a majority of others in the circumstance; the likelihood of helping behavior depends in large degree on the numbers of other people present and the degree of familiarity of the subject with the other subjects in the situation (Asch 1951; Latané and Rodin 1969); social preferences are driven by subliminal smells (Li et al. 2007); one's name can play a startlingly large role in one's important life choices, and so on (Pelham et al. 2002). Collectively, such work is known as *situationist social psychology*, or *situationism.*

The general lesson of situationism is that we underestimate the influence of the situation and we overestimate the influence of purportedly fixed features of the agent. Crucially, the "situational inputs" typically operate without the awareness of the agent. Seemingly inconsequential —and deliberatively irrelevant— features of the context or situation predict and explain behavior, suggesting that our agency is somewhat less than we presume. Indeed, when agents are asked about the relevance of those apparently innocuous factors in the situation, the usual reply is outright denial or dismissal of their relevance to the subject's deliberation and decision. Thus, contemporary psychological science threatens the plausibility of Reasons accounts by showing that the basis of our actions are disconnected from our

assessments of what we have reason to do.[5]

While particular experiments may be subject to principled dispute, the general lesson—that we frequently underestimate the causal role of apparently irrelevant features of contexts on our behavior, both prospectively and retrospectively—has considerable support (Doris 2002, p. 12-13).[6] There are ongoing disputes about the precise implications of situationism, and in particular, what this data shows about the causal role of personality and what implications this might have philosophical theories of virtue (Doris 1998; Harman 1999; Kamtekar 2004; Merritt 2000; Sabini et al. 2001). In the present context, however, these concerns can be bracketed. What follows does not obviously depend on the nature of personality or character traits, and so whatever the status of those debates we have reason to worry about situationism's implications for Reasons views.

The situationist threat operates on two dimensions. On the one hand, it may threaten our "pre-theoretical" or *folk* view of free will. On the other hand, to the extent to which situationism suggests we lack powers of agency that are appealed to on philosophical accounts, it threatens our philosophical theories of free will. In what follows I focus on the *philosophical* threat, and in particular, the threat to Reasons accounts. Whatever we say about folk judgments of freedom and responsibility, what is crucial here is what our best theory ought to say about free will, all things considered.

None of this implies that situationism's significance for ordinary beliefs is altogether irrelevant for philosophical accounts. On the contrary: Reasons accounts are partly motivated by their coherence with ordinary

---

5 In ordinary discourse, to say something "is a threat" is to say something ambiguous. It can mean either the *appearance* of some risk, or it can indicate the *actuality* of risk or jeopardy, where this latter thing is meant in some non-epistemic way. When my kids and I pile out of the minivan and into our front yard, I strongly suspect our neighbors regard us as a threat to the calm of the neighborhood. Nevertheless, there are some days on which the threat is only apparent. Sometimes we are simply too tied to yell and make the usual ruckus. As I will use the phrase *the situationist threat,* it is meant to be neutral between the appearance and actuality of jeopardy. Some apparent threats will prove to be only apparent, and others will be more and less actual to different degrees.

6 We must be careful, though, not to *over*claim what the body of literature gets us. Although it is somewhat improbable that one could do so, it is possible that one could generate an alternative explanation that (1) is consistent with the data but that (2) does not have the implication that we are subject to situational effects that we misidentify or fail to recognize. This would be a genuine problem for the situationist program.

judgments. If ordinary judgments turn out to be at odds with the scientific picture of our agency, this undercuts some of the motivation for accepting a Reasons theory. A Reasons theorist might be willing to sever the account's appeal to ordinary judgments, but if so, then something needs to be said about the basis of such accounts. For conventional Reasons theorists, however, situationism presents an unappealing dilemma: we can either downgrade our confidence in Reasons theories (in light of the threat of situationism) or we can disconnect our theories from ordinary judgments and downgrade our confidence in our ordinary judgments of responsibility.

So, the dual threat situationism presents to common sense and philosophical theorizing does not easily disentangle. Nevertheless, my focus is primarily on the philosophical threat, whatever its larger implications for our ordinary thinking and its consequent ramifications for theorizing.[7]

3. Situationism, Irrationality, and Bypassing

Situationism might impugn our rationality in at least two ways. It might show that those psychological processes that bring about action are irrational. Alternately, it could show that what rationality we have is too shallow to constitute freedom. I explore these possibilities in turn.

Let's start with a non-human case of impugned rationality.

The common digger wasp can act in some intriguingly complex ways. Consider its behavior in providing food for its eggs. It will drag a stung cricket to the threshold of its burrow, release the cricket, enter the burrow for a moment (presumably to look things over), and then return to the threshold of

---

7 For what it is worth, I suspect that the one source of the perception of a situationist threat is traceable to an overly simple description of the phenomena. Recall the data in *Phone Booth*. We might be tempted to suppose that what it shows is that agents are a site upon which the causal levers of the world operate, if we describe it as a case where "the dime makes the subject help." Such descriptions obscure something important: the fact of the agent's psychology and its mediating role between situation and action. The more accurate description of *Phone Booth* seems to be this: the situation influences (say) the agent's mood, which affects what the agent does. Once we say this, however, we are appealing to the role of some psychological elements that presumably (at least sometimes) constitute inputs to and/or elements of the agent's active, conscious self. If we focus on this fact, we do not so easily lose the sense of the subject's agency. A coarse-grained description of situationist effects may thus sometimes imply a form of fatalism that bypasses the agent and his or her psychology. More on a related issue in §4.

the burrow to pull the cricket in. Here is the surprising thing: if you move the cricket more than a few inches from the threshold of the burrow when the wasp enters its burrow for the first time without the cricket, the wasp will "re-boot" the process: moving the cricket closer, dropping it, checking out the burrow, and returning outside to get the cricket. Surprisingly, the wasp will do this *every time the cricket is moved*. Again, and again, and again, and again, if necessary. The wasp never does the obvious thing of pulling the cricket into the burrow straightaway. Douglas Hofstadter calls this *sphexishness*, or the property of being mechanical and stupid in the way suggested by the behavior of the digger wasp (*Sphex ichneumoneus*) (Dennett 1984, p. 10-11).

Now consider the human case. Perhaps situationism shows that we are sphexish. We think of ourselves as complicated, generally rational creatures that ordinarily act in response to our best assessments of reasons. What situationism shows, perhaps, is that we are not like that at all. Instead, our agency is revealed as blind instinct, perhaps masked by high level confabulation (i.e., the manufacturing of sincere but *ad hoc* explanations of the sources of our action).[8] If one thinks of instinct as paradigmatically opposed to free will, then we have a compact explanation for why situationism threatens free will: it shows we are instinctual and not free.

Recall, however, that the present issue is not whether our naive, pre-theorized views of the self are threatened by situationism. What is at stake is whether a Reasons theory of free will should be threatened by situationism. In this context, it is much harder to make out why it should matter that some of our rational capacities reduce to the functioning of lower level "instinctual" mental operations. One way to put the point is that it is simply a mistake to assume that instinct is *necessarily* opposed to rationality. To be sure, some behavior we label 'instinctive' might be, in some cases, irrational by nearly any measure. Still, there are cases where instinctive behavior is ordinarily rational in a straightforwardly instrumental sense. In a wide range of circumstances, our "instincts" (to breath, or to jerk our hands away from burning sensations, to socialize with other humans, and so on) are paradigms of rational behavior. What we call 'instinct' is (at least sometimes) Mother Nature's way of encoding a kind of bounded rationality into the basic mechanisms of the

---

8 This idea plays a particularly prominent role in the work of Daniel Wegner and proponents of recent research on the automaticity of mental processes (e.g., John Bargh). See § 4.

creature. We view it as *mere* instinct only when we find the limits of that rationality, or when it conflicts with our higher-order aims.

On this picture, the fact that we are sphex-ish (in the sense of having an instinctual base for rational behaviors) does not threaten our freedom. Instinctual behaviors can be irrational, especially when they operate under conditions different than they were presumably acquired, or when they come into conflict with some privileged element of the psychic economy.[9] However, neither the fact of instinctual behavior, nor the possibility of a reduction of our complex behavior to more basic, less globally rational elements shows that we cannot respond to reasons any more than pockets of local irrationality in the wasp would show us anything about the wasp's rationality under normal conditions. Our rationality is, like the wasp's, plausibly limited. Such limitations, however, do not constitute global irrationality.

The line of response suggested here—construing instinctual, sub-agential mechanisms with bounded rationality as partly constitutive of our general rationality—might suggest a different sort of worry. On this alternative worry, what situationism shows is not that there is no sense to be made of the idea that we might be rational, but rather that what rationality we have is a function of processes that *bypass* our agency. That is, under ordinary circumstances our behavior might be rational or not, but what drives that behavior does not involve a contribution of active, deliberating agency. To the extent to which a Reasons account requires that the seat of reasons-responsive agency be a conscious, deliberative self, situationism will threaten this picture of free will.

If situationism could show this much, this might indeed constitute a threat that Reasons theorists should take seriously. However, there are good reasons to doubt that situationism has shown this much. Even if our agency were oftentimes bypassed, this would not obviously undermine all attributions of free will. At least in ordinary practice, we do not hold that agents must *always* exercise their conscious, active agency in order to be responsible for the outcome. Negligence, for example, is typically treated as a case of responsibility where the failure to act need not be the product of an intentional or conscious

---

9 On some views, the privileged psychological element could include things such as the agent's conscious deliberative judgments or some maximally coherent arrangement of the agent's desires and beliefs.

choice not to act.[10] And, in a range of cases, we seem perfectly willing to regard ourselves and others as responsible for actions that arrive unexpectedly, but whose arrival we regard as telling us something about where we stand. One's enthusiasm (or lack thereof) in response to, for example, a marriage proposal, a job offer, or an invitation to a weekend away can be unexpected. Nevertheless, we can (and oftentimes do) regard those reactions as privileged and the emotional bedrock on which the praiseworthiness and blameworthiness of subsequent action are anchored (Arpaly 2003). So, the mere fact that our active agency is sometimes bypassed does not obviously show we lack the freedom sufficient for moral responsibility.

Reactions that bypass our conscious, deliberative selves could be a problem if, for example, the bypassing mechanisms were themselves never or rarely responsive to reason. However, this possibility suffers from the same basic problem that plagued the deterministic gloss on situationism: the evidential base does not support so sweeping a generalization. Unless we receive compelling evidence for the possibility—evidence that extends far beyond the idea that situation play a larger, oftentimes puzzling, role in action than we ordinarily acknowledge—the mere fact of our actions bypassing or override conscious deliberation is not, in itself, problematic for a Reasons theory.[11]

Still, there seems to be a lurking difficulty for the Reasons theorist. If we concede that experimental data can show that there are conditions under which we are irrational, we might wonder what the point is at which we become too irrational for our ordinary practices of moralized praise and blame to retain their integrity. Too much irrationality too often might mean that we cannot often enough assume people satisfy the conditions of rationality that hold on free will for us to go on as we did before.

---

10 Negligence is a particularly difficult to account for aspect of moral responsibility, so perhaps this is not so telling. Matt King has recently argued against treating negligence as a case of responsibility precisely because it lacks the structure of more paradigmatic cases of responsibility (King 2009)

11 The bypass threat might work in a different way. Perhaps the worry is not that our conscious deliberative agency is sometimes trumped by our emotions. Perhaps the picture is, instead, that our active, deliberative agency *never* plays a role in deciding what we do. Perhaps situationist data suggests that our active, mental life is a kind of sham, obscuring the operation of sub-agential processes beyond our awareness. I am dubious, but for the moment I will bracket this concern, returning to it when I discuss the automaticity literature.

To resolve this matter, we require two things: (1) more detailed experimental data than we currently have, and (2) an account of what sort of rational powers we need for free will and moral responsibility. In their absence, it is difficult to evaluate whether the frequency and degree of irrationality we ordinarily exhibit undermine free will. I cannot provide the former, but in a bit I will attempt to provide a sketch of the latter: it is, I think, enough to blunt some of the worry, even if it cannot eradicate it altogether First, though, I want to consider one further issue that might be taken to supplement the basic worry generated by situationism.

4. An Automaticity Threat?

A suitably informed interlocutor might contend that even if situationism lacks the resources to show that we lack free will, other work in psychology can do so. In recent years a fertile research program has sprung up around detailing the scope of fast, nonconscious determinants of action and preferences and their mechanisms of operation. This work is usually thought of as describing the *automaticity* of human action. Automatic processes, if sufficiently irrational and pervasive, would presumably show that we are not often enough responsive to reasons.

As John Kihlstrom characterizes it, automatic processes have four features:

> 1. Inevitable evocation: Automatic processes are inevitably engaged by the appearance of specific environmental stimuli, regardless of the person's conscious intentions, deployment of attention, or mental set.
> 2. Incorrigible completion: Once evoked, they run to completion in a ballistic fashion, regardless of the person's attempt to control them.
> 3. Efficient execution: Automatic processes are effortless, in that they consume no attentional resources.
> 4. Parallel processing: Automatic processes do not interfere with, and are not subject to interference by, other ongoing processes—except when they compete with these processes for input or output channels, as in the Stroop effect (Kihlstrom 2008).

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 (Bargh and Ferguson 2000).

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.

The work on automaticity raises some interesting questions of its own, and it merits a much more substantial reply than I will give to it here. Still, because concerns about automaticity interlock with the situationist threat, it may be useful to sketch what sorts of things the Reasons theorist might say in reply to automaticity worries.

First, it is hardly clear how much of our mental life is automatic in the way defined at the start of this section (Kihlstrom 2008). There are ongoing disputes about this issue among psychologists, and the dust is not yet settled, especially with respect to the matter of the ubiquitousness of diversity of automatic processes and the extent to which ballistic processes are immune to deliberate modification. Notice, though, that as long as the formation of conscious intentions has work to do in the constraining of courses of action and the assignments of weights to elements of deliberation, and as long as those processes can respond to reasons, there seems to be room for a picture of agency that characterizes the responsibility-relevant notion of freedom or control in terms of rational agency.

A critic might object that this picture of intentional control is precisely what psychologists (from the earlier work of Benjamin Libet to more recent work by Daniel Wegner and John Bargh) have been denying. Despite the considerable attention this work has generated, some of the most dramatic claims—for example, that conscious intentions do no work in action production—have been subject to trenchant criticism on both conceptual *and* empirical grounds (Mele 2009; Nahmias 2002; Nahmias 2007; Bayne 2006). So, the Reasons theorist will surely be quick to note that matters are not obviously settled in the skeptic's favor.

Moreover, in response to threats from automatic processes, many of the same replies as were offered against situationism are available. First, that a process is automatic does not mean it is necessarily irrational. What matters is whether we are appropriately responding to reasons, regardless of whether we are thinking of them as such. So, the free will skeptic would need to show both

that our automatic processes are ubiquitous *and* overwhelmingly irrational in ordinary action. Second, that sub-agential automatic processes can partly constitute our agency does not mean that those automatic processes are necessarily not attributable to us. It may well make sense to regard a good many automatic processes as attributable to us, depending on the details. Third, automatic does not mean uncontrolled, especially if the agent embraces the presences of those automatic processes, has knowingly inculcated them, or they appropriately contribute to a disposition, aim, or practice that is valuable to the agent.

This is all too quick, of course. Sorting out the details will require painstaking work I will not attempt to pursue here. Still, considerations along the lines of those I have sketched will surely be part of what a Reasons theorists will say in reply to more detailed objections derived from research on automaticity. Here, as elsewhere, the task of reading off philosophical ramifications from empirical work is a messy business.

Presumably, some automatic processes will not be suitably responsive to reasons—moral or otherwise—in a range of comparatively ordinary circumstances of action. To that extent, what philosophers have to learn from psychology and related sciences are the contours of ordinary dispositions to rationally respond to pertinent considerations. The degree to which the empirical data raises philosophical problems, however, is unlikely to be settled in the philosopher's armchair or the scientist's lab, for (as I argue in the next few sections) the issues are fundamentally both empirical *and* philosophical.

5. Giving up some assumptions?

Situationism does not show that we are always irrational, or that situational forces always bypass our agency. So, situationism does not threaten for those reasons. Still, a critic might note, even if we are sometimes (perhaps even regularly) as rational as we can realistically hope to be, our rational, moral natures are very fragile and bounded. The critic might charge that *this* is the real situationist threat.

That seems right to me. In order to better address this criticism, however, I think we must recast Reasons accounts, abandoning some suppositions that are usually folded into such accounts. Let me explain.

Many accounts of free will are implicitly committed to something I shall call *atomism.* Atomism is the view that free will is a non-relational property

of agents; it is characterizable in isolation from broader social and physical contexts. An *atomist* (in the present sense) holds that whether a given agent has free will and/or is capable of being morally responsible can, at least in principle, be determined simply by reading off the properties of just the agent. Atomistic theories provide characterizations of free will or responsible agency that do not appeal to relational properties, such as the normative relations of the agent to institutions or collectives.

Atomism is often coupled with a view that there is only one natural power or arrangement of agential features that constitutes free will or the control condition on moral responsibility. This is a *monistic* view of the ontology of free will. Monistic views include those accounts that hold that free will is the conditional ability to act on a counterfactual desire, should one want to. Identificationist accounts, which hold that free will is had only when the agent identifies with a special psychological element, (a desire, a value, an intention, etc.) are also monistic. So are libertarian accounts, on which one acts freely only when one acts in a specific non-deterministic fashion. In contrast, non-monistic (or pluralistic) accounts hold that there are multiple agential structures or combinations of powers that constitute the control or freedom required for moral responsibility.

If we assume that the freedom or control implicated in assessments of moral responsibility is a single, unified capacity that relies on a particular cross-situationally stable mechanism, then the sciences of the mind will be threatening to these accounts. The situation-dependent nature of our capacities seems to be perhaps the most compelling claim of situationist research. Consequently, the implicit picture of our natural capacities invoked by going philosophical theories—an atomistic, monistic picture—looks to be just plain false.

Psychological research suggests that what appears to us as a general capacity of reasons-responsiveness is really a cluster of more specific, ecologically limited capacities indexed to particular circumstances. Consequently, what powers we have are *not* had independently of situations. What capacity we have for responding to reasons is not some single thing, some fixed structure or cross-situationally stable faculty.

Importantly, degradation of our more particular capacities can be quite localized and context-specific. Consider the literature on "stereotype threat" or "social identity threat." What Steele, Aaronson, and their colleges have found

is that performance in a wide range of mental and physical activities is subject to degradation in light of subjects perceiving that there is some possibility of their being evaluated in terms of a negative stereotype (Aronson et al. 1999; Steele et al. 2002). So, for example, when there is a background assumption that women and blacks do less well then white men at math, the performance of women and blacks on math exams—a task that plausibly involves a species rationality, if anything does—will drop when the exam is presented as testing native ability. These startling results disappear when the threat is removed, as when, for example, the exam is presented as testing cognitive processes and not purportedly native ability. One can do the same thing to white males, by priming them with information about their stereotypically poor performance on math tests when compared to their Asian counterparts. When the threatening comparison is made salient to subjects, performance drops. When the threatening comparison is taken away, and the exam is explicitly presented as *not* susceptible to such bias, scores rise for the populations ordinarily susceptible to the threat.

Remarkably, these results generalize to a variety of more and less cognitive domains, including physical performance (Steele et al. 2002).[12] Indeed, the more general thesis, that the environment can degrade our cognitive capacities in domain-specific ways, has considerable support (Doris and Murphy 2007). One could resist the general lesson by arguing that (perhaps) there is a basic underlying capacity that is stable, and perception (whether conscious or not) of stereotypes affects the ease with those capacities are exercised. Notice, though, that this just pushes the problem with atomistic views back a level. Even if our basic capacities are stable across contexts, our abilities to exercise them vary by circumstance and this suggests that our situation-indexed capacities vary considerably.

Given that free will is a fundamentally practical capacity—it is tied to action, which always occurs in a circumstance—then the characterization of our freedom independent of circumstance looks like a vain aspiration. What we need to know is whether we have a capacity relevant for action (and, on the present interpretation, responsible action)— this requires an account of free will that is sensitive to the role of the situation. An atomistic account cannot

---

12 One remarkable result from that study: women shown TV commercials in stereotypically unintelligent roles before an exam led to worse performance on math tests (393).

hope to provide this, so we must build our account with different assumptions.

There are various ways the conventional Reasons theorist might attempt to rehabilitate atomism and monism. In what follows, however, I explore what possibilities are available to us if we retain an interest in a Reasons account but pursue it without the assumptions of atomism and monism.

6. Reasons-responsiveness reconceived

Situationism presses us to acknowledge that our reasons sensitive capacities are importantly dependent on the environment in which those capacities operate, and that the cross-situational robustness of our reasons-responsive agency is a sham. At its core, the idea is intuitive enough— the power of a seed to grow a tree is only a power it has in some contexts and not others. The challenge is to remember that this is true of persons, too, and that this generates the corresponding need to appreciate the circumstances that structure our powers.[13]

In this section, my goal is to provide an account that is (1) consistent with a broadly Reasons approach, (2) free of the supposition of atomism and monism about the involved agential powers, and (3) compatible with a wide range of plausible theories of normative ethics. So, the account I offer is one where the characteristic basic structure of responsible agency is to be understood as a variably-constituted capacity to recognize or detect moral considerations in the relevant circumstances, and to appropriately govern one's conduct in light of them.

It may help to start by contrasting the account to a more familiar approach. Consider the traditional "conditional analysis" of classical compatibilism. On this approach, to say that an agent has the capacity to do otherwise is to attribute a conditional power (or, perhaps, a conditional analysis of a categorical power): were one to decide to do otherwise, one would do otherwise.

---

13 Some social psychologists have contended that the degree to which populations emphasize indivdual vs. situation in explanation and prediction varies across cultures (Nisbett 2003). Recently, the idea that circumstances structure decision-making in subtle and under-appreciated ways has recently received popular attention because of the visibility of (Thaler and Sunstein 2009). The present challenge is to provide an characterization of what the responsibility-relevant notion of control comes to given that our decisions are vulnerable to "nudges" of the sort they describe.

The traditional conditional analysis was elegant and problematic in equal measure. In recent years there have been a number of intriguing attempts to resurrect the general approach.[14] Whatever the virtues of those accounts, the picture I am offering is rather different. On the account I favor, the powers that constitute free will are precisely those that are required for a sufficiently good arrangement of praising and blaming practices, one that has as its aim the cultivation of our recognizing and appropriately responding to moral considerations.[15]

Let us start with a relatively uncontroversial characterization of the terrain, given the presumption that free will is the capacity distinctive in virtue of which agents can be morally responsible. On this picture, we can say this:

> For an agent *S* to be responsible for some act token *A* in context *C* requires that *S* is a responsible agent and the action is morally praiseworthy or morally blameworthy.

The present schema invokes several technical notions: the idea of a *responsible agent*, and an account of what it is for an action to be *morally praiseworthy* and *morally blameworthy*. I will leave the latter two notions unanalyzed, focusing on the implications of abandoning the standard atomistic and monistic model of responsible agency and its capacities.

Here is how I think the Reasons theorist should characterize responsible agency, and by extension, free will:

> An agent S is a responsible agent with respect to considerations of type M in circumstances C if S possesses a suite of basic agential capacities implicated in effective self-directed agency (including, for example, beliefs, desires, intentions, instrumental reasoning, and generally reliable beliefs about the world and the consequences of action) and is also possessed of the relevant capacity for (A) detection of suitable moral considerations M in C and (B) self-governance with respect to M in C. Conditions (A) and

---

14 For a useful overview of the difficulties faced by the classical conditional analysis, see (Kane 1996). For a critical discussion of more recent approaches in this vein, see (Clarke 2009)

15 Much of the machinery I introduce to explicate this idea can, I think, be paired with a different conception of the normative aim for moral responsibility; the specific powers identified will presumably be somewhat different, but the basic approach is amenable to different conceptions of the organizing normative structure to the responsibility system. I leave it to other to show how that might go.

(B) are to be understood in the following ways:

(A) the capacity for detection of the relevant moral considerations obtains when:

(i) S actually detects moral considerations of type M in C that are pertinent to actions available to S or

(ii) in those possible worlds where S is in a context relevantly similar to C, and moral considerations of type M are present in those contexts, in a suitable proportion of those worlds S successfully detects those considerations.

(B) the capacity for volitional control, or self-governance with respect to the relevant moral considerations M in circumstances C obtains when either

(i) S is, in light of awareness of M in C, motivated to accordingly pursue courses of action for which M counts in favor, and to avoid courses of action disfavored by M or

(ii) when S is not so motivated, in a suitable proportion of those worlds where S is in a context relevantly similar to C

(a) S detects moral considerations of type M, and

(b) in virtue of detecting M considerations, S acquires the motivation to act accordingly, and

(c) S successfully acts accordingly.

And, the notions of suitability and relevant similarity invoked in Aii and Bii are given by the standards an ideal, fully-informed, rational, observer in the actual world would select as at least co-optimal for the cultivation of our moral reasons-responsive agency, holding fixed a range of general facts about our current customary psychologies, the cultural and social circumstances of our agency, our interest in resisting counterfactuals we regard as deliberatively irrelevant, and given the existence of genuine moral considerations, and the need of agents to internalize norms of action for moral considerations at a level of granularity that is useful in ordinary deliberative and practical circumstances. Lastly, the ideal observer's determination is structured by the following ordering of preferences:

(1) that agents recognize moral considerations and govern themselves accordingly in ordinary contexts of action in the actual world

(2) that agents have a wider rather than narrower range of contexts of action and deliberation in which agents recognize and respond to moral considerations.

So, free will is a composite of conditions A and B. In turn, A and B are subject to varied ways of being constituted in the natural world. It is a picture

of free will that can be had without a commitment to atomism and monism of the sort that the contemporary sciences of the mind impugn.

Before exploring the virtues of this account, some clarification is in order. First, the above characterizations make use of the language of possible worlds as a convention. The involved locutions (e.g., "in those worlds") is not meant to commit us to a particular conception of possibility as referring to concrete particulars. Second, the possibilities invoked in the above account are—by design—to be understood as constituting the *responsibility-relevant* capacities of agents. These capacities will ordinarily be distinct from the "basic abilities," or the intrinsic dispositions of agents.[16] Instead, they are higher order characterizations picked out because of their relevance to the cultivation and refinement of those forms of agency that recognize and respond accordingly to moral considerations of the sort we are likely to encounter in the world. This is a picture on which the relevant metaphysics of our powers is determined not by the physical structures to which our agency may reduce, but instead by the roles that various collections of our powers play in our shared, normatively structured lives. Third, the account is neutral on the nature of moral considerations. Moral considerations presumably depend on the nature of right and wrong action and facts about the circumstances in which an agent is considering what to do.[17]

So, where does all of this get us? Characterizing the capacities that constitute free will as somewhat loosely connected to our intrinsic dispositions allows us to reconcile the picture of our agency recommended by the psychological sciences without abandoning the conviction that our judgments and practices of moral responsibility have genuine normative structure to them. We are laden with cognitive and sub-cognitive mechanisms that (however ecologically bounded) sometimes can and do operate rationally. There are surely times when our autonomic, non-conscious responses to features of the world come to hijack our conscious plans. When this occurs, sometimes it will

16 I borrow the term "basic abilities" from John Perry, although my usage is, I think, a bit different (Perry 2010).

17 I favor the language of moral *considerations* (as opposed to moral *reasons*) only because talk of reasons sometimes is taken to imply a commitment to something like an autonomous faculty that properly operates independently of the effects of affect. There is nothing in my account that is intended to exclude the possibility that affect and emotion, in both the deliberating agent and in those with whom the agent interacts, play constitutive roles in the ontology of moral considerations.

mean we are not responding to reasons. Other times what it will mean that we are responding to reasons, just not the reasons our conscious, deliberative selves are aware of, or hoping to guide action. Still, this fact does not mean that we are incapable of recognizing and responding to reasons, even moral reasons. The facts concerning our unexercised capacities, at least as they pertain to assessments of the responsibility-relevant notion of control, depend on counterfactuals structured by normative considerations.[18]

A distinctive feature of the account is that foregrounds a pluralist epistemology of moral considerations. Recognition or sensitivity to moral considerations is *not* a unified phenomenon, relying on a single faculty or mechanism. Moral considerations may be constituted by or generated from as diverse things as affective states, propositional content, situational awareness, and so on. Consequently, the corresponding epistemic mechanisms for apprehending these considerations will presumably be diverse as well.[19] Moreover, the present picture does not hold that sensitivity to moral considerations must be conscious, or that the agent must recognize a moral consideration *qua* moral consideration for it to count as such. An agent may be moved by moral considerations without consciously recognizing those considerations and without conceiving of them as *moral* considerations.[20]

Another notable feature of the account is that it makes free will variably had in the same individual. That is, the range of moral considerations an agent recognizes in some or another context or circumstance will vary. In some circumstances agents will be capable of recognizing a wide range of moral considerations. In other circumstances those sensitivities may be narrower or

---

18 Notice that even if the skeptic is right that we are very often not suitably responsive to moral considerations, the present account suggests that there may yet be some reason for optimism, at least to the extent to which we can enhance our capacities and expand the domains in which they are effective.

19 It would be surprising if the epistemic mechanisms were the same for recognizing such diverse things as that someone is in emotional pain, that other persons are ends in themselves, and that one should not delay in getting one's paper to one's commentator. The cataloging of the varied epistemic mechanisms of moral considerations will require empirical work informed by a more general theory of moral considerations, but there is already good evidence to suggest that there are diverse neurocognitive mechanisms involved in moral judgments (Nichols 2004; Moll et al. 2005).

20 Huck Finn may be like this, when he helps his friend Jim escape from slavery. For an insightful discussion of this case, and the virtues of *de re* reasons responsiveness, see (Arpaly 2003).

even absent. When they are absent, or when they dip beneath a minimal threshold, the agent ceases to be a responsible agent, *in that context*. We need not suppose that if someone is a responsible agent at a given time and context, that he or she possesses that form of agency at *all* times across *all* contexts. In some contexts I will be a responsible agent, and in others not. Those might not be the same contexts in which you are a responsible agent.

When we have reason to believe that particular agents do not have the relevant sensitivities or volitional capacities in place, we do not believe that they are genuinely responsible, even if we think that in other circumstances the agent does count as responsible. We may, however, respond in responsibility-characteristic ways with an eye towards getting the agent to be genuinely responsible in that or related contexts. Or, we may withhold such treatment altogether if we take such acculturation to be pointless, not worth the effort, or impossible to achieve in the time available to us.[21] However, we can understand a good deal about the normative structure of moral responsibility if we think of it as modestly teleological, aiming at the development of morally-responsive self-control and the expansion of contexts in which it is effective.

This limited teleology is perhaps most visible in the way a good deal of child-rearing and other processes of cultural inculcation are bent to the task of expanding the range of contexts in which we recognize and rightly respond to moral considerations (Vargas 2010). By the time we become adults, praise and blame have comparatively little effect on our internalizing norms, for we oftentimes have come to develop habits of thought and action that deflect the force of moral blame directed at us. Still, the propriety of our judgments turns on facts about whether we are capable of recognizing and appropriately responding the relevant moral considerations in play.

7. Re-assessing the situationist threat

All of this is well and good, one might reply, but how does this help us address

---

21 Cases of this latter sort can occur when one visits (or is visited by) a member of a largely alien culture. In such cases, we (at least in the West, currently) tend toward tolerance of behavior we would ordinarily regard as blameworthy precisely because of the conviction that the other party operates out of an ignorance that precludes apprehension of the suitable moral considerations. As George Fourlas pointed out to me, in recently popular culture, this phenomenon has been exploited to substantial and controversial comedic effect by comedian Sasha Baron Cohen.

the situationist threat? Too see how, we can revisit the situationist experiments mentioned at the outset of the paper.

In retrospect, the Isen and Levin's *Phone Booth* seems unproblematic. On the present account, the fact that a situation might radically alter our disposition to respond to reasons to help is neither puzzling nor especially troubling. As mentioned before, the natural explanation here seems to be the effects of the dime on mood. There is consensus among psychologists that mood affects helping behavior (Weyant 1978). In this particular case, there is nothing to suggest that the agent has been robbed of the capacities that constitute free will. The basic capacities we have reason to be worried about in ascribing responsibility appear to be intact, the influence of the situation is benign (i.e., enhancing willingness to help others), and anyway, the helping may well be supererogatory.[22]

Situations may influence mood, and mood may affect the likelihood of some or another resultant action, but those influences (unless radically debilitating) do not usually change the presence or absence of the capacities that constitute free will. What the present account helps to explain is why the mere fact of a change in the likelihood of some action (e.g., because of a change in salience of some fact in the agent's deliberations or the effect of mood)—or even a fundamental change in capacity to do otherwise in the relevant sense—does not automatically entail that the agent lacks free will, however counterintuitive that might initially strike us (Talbert 2009; Vargas and Nichols 2007). The higher level capacities that are required for moral responsibility are not necessarily disrupted by such changes.

There is some complexity to the way mood affects behavior, and it raises a potential difficulty for the present account. Positive moods generally increase helping behavior, in contrast to neutral moods. However, the effect of *negative* moods on helping behavior is varied. It is especially sensitive to the cost to the agent and the apparent benefit generated by the helping in some interesting ways. In cases where helping is of low cost to the agent but of plausibly high benefit (e.g.volunteering time to fundraise by sitting at a

---

22 The supererogatory nature of helping can be important if, for example, one is worried about the *not* helping condition, and how infrequently people help strangers with minor problems. Perhaps one more global issue here is simply how rare it is that we act on moral considerations, whether because of failures of perception or motivation. I return to this issue, at least in part, at the end.

donation desk for the American Cancer society), negative moods actually *increases* helping behavior over neutral moods. However, in cases where the benefits are low and the costs to the agent are high (going door-to-door to raise funds for the Little League), negative moods tends to mildly suppress helping behavior (Weyant 1978).[23] In cases where both the benefits and costs are matched high or matched low, negative moods have no effect over neutral states.

These data may suggest a problem for the present account. Perhaps what the mood data show is that the agent is not being driven by reasons so much as an impulse to maintain equilibrium in moods. According to this line of objection, helping is merely an instrumental means to eliminate the bad mood, albeit one that is structure by the payoffs and challenges of doing so. If this is so, however, then it appears that agents in bad moods do not seem to be helping for good reasons, or even moral reasons at all. Consequently, the present account seems to have made no headway against the threat that experimental psychology presents to Reasons accounts.[24]

I agree that the role of mood in agents is complex. Still, I think the challenge can be met. As an initial move, we must be careful not to presume that affective states and moral reasons are always divorced. Plausibly, moral considerations will be at least partly constituted by an agent's affective states. Moreover, an agent's affective states will play a complex role in the detection of what moral considerations there are. So, what mood data might show is not that agents in negative moods do not help for good reasons or for no moral reasons at all, but rather that being in negative moods can make one aware of, or responsive to, particular kinds of moral reasons.[25] Commiseration and sympathy are quite plausibly vehicles by which the structure of morality becomes integrated with our psychology. And, as far as I can tell, nothing in the mood literature rules this possibility out. Indeed, what we may yet have reason to conclude is that the mechanisms of mood equilibrium are some of the main mechanisms of sympathy and commiseration. To note their activity

---

23 For a helpful discussion of this literature, and its significance for philosophical work on moral psychology, see (Miller 2009b; Miller 2009a).

24 Thanks to Christian Miller for putting this concern to me.

25 Note that none of this requires that the agent conceive of the reasons as moral reasons. As noted earlier, the relevant notion of responding to moral reasons is, borrowing Arpaly's terminology, responsiveness *de re*—not *de dicto*.

would thus not undermine a Reasons picture so much as it would explain its mechanisms.[26] If all this is correct, the proposed account can usefully guide our thinking about *Phone Booth* and related examples.

Now consider *Samaritan*. What this experiment appears to show is that increased time-pressure decreases helping behavior. Non-helping behavior is, presumably, compatible with free will. An agent might decide to not help. Or, depending on how that subject understands the situation, he or she might justifiably conclude that helping is supererogatory. So, decreased helping behavior is not direct evidence for absence of free will. Still, perhaps among some agents in *Samaritan* suffered a loss of free will.

Here are two ways that might have happened. First, if what happened in *Samaritan* is that time-pressure radically reduced the ability of agents to recognize that someone else needs help (which is what at least some of the subjects reported), then this sort of situational effect can indeed undermine free will precisely by degrading an agent's capacity to recognize moral considerations. So, perhaps some *Samaritan* subjects were like this. A second way to free will in *Samaritan*-like circumstances could be when time-pressure sufficiently undermines the ability of the agent to act on perceived pro-helping considerations.

A natural question here is how much loss of ability constitutes loss of capacity. Here, we can appeal to the account given above, but it does not give us a bright line. At best, it gives us some resources for what sorts of things to look for (e.g., what data do we have about how much time-pressure, if any, saps ordinary motivational efficacy of recognized moral considerations). Some of these issues are quasi-empirical matters for which more research into the limits of human agency is required. Still, in the ordinary cases of subjects in *Samaritan*, it seems that we can say this: if one did not see the "injured" person, then one is not responsible. Matters are more complicated if one did see the "injured" person and thought he or she needed help, and the agent thought him- or herself capable of helping without undue risk (the *Samaritan* study did

---

26 What would undermine the Reasons approach? Here's one possibility amenable to empirical data: If mood data showed that people were driven to increased helping behavior when they ought not (e.g., if the only way to help would be to do something *really* immoral), this would suggest that at least in those cases mood effects were indeed disabling or bypassing the relevant moral considerations-sensitive capacities. But in mood-mediated cases, such behavior is rarely ballistic in this way.

not distinguish between agents who had and lacked these convictions). In such circumstances, I am inclined to think that one could avoid blameworthiness only if, roughly, time- or other situational pressure was sufficiently great that most persons in such circumstances would be incapable of bringing themselves to help. It is, as far as I know, an open question whether there is any empirical data that speaks to this question, one that folds in the agent's understanding of the situation. So, I think, the *Samaritan* data does not give us a unified reason to infer a general lack of free will in time-pressure cases; sometimes time-pressure may reduce helping behavior because of any number of reasons, only some of which are free will disrupting.

Finally, let us reconsider the Milgram *Obedience* cases. Here, there is some reason to doubt that subjects in the described situations retain the responsibility-relevant capacities. At least in very general terms, *Obedience*-like cases are precisely ones in which agents are in unusual environments and/or subject to unusual pressures. Plausibly, they are situations that stress the ordinary capacities for responding to reasons, or they invoke novel cognitive and emotional processes in agents. This shift in situation, and the corresponding psychological mechanisms that are invoked, may decrease the likelihood that an ordinary agent will have the responsibility-relevant capacities. We check, roughly, by asking whether in a significant number of deliberatively relevant circumstances the evaluated agent would fail to appropriately respond to the relevant moral considerations. *Ceteris paribus*, the higher our estimation of the failure rate in a given agent, the more reason we have to doubt that the agent possesses the capacity required for moral responsibility. Still, in *Obedience*-like situations, agents are not necessarily globally insensitive to moral considerations or even insensitive to only relevant moral considerations. Some may well be suitably sensitive to some or all of the moral considerations in play. (Indeed, some subjects did resist some of the available abhorrent courses of action with greater and lesser degrees of success.) So, there is a threshold issue here, and in some cases it will be comparatively unclear to us what an ideal observer would say about a case thus described.

8. What about active, self-aware agency?

Before concluding, it may be useful to remark on what role, if any, the present picture leaves for agency of the active, self-aware variety. Suppose that we accept that some of our reasons-detecting processes are conscious and others

are not, and that some move from conscious to unconscious (or in the opposite direction) through sufficient attention or practice. The outputs of these varied mechanisms will sometimes converge, and other times conflict. What role there is for active, conscious, deliberative agency is crucial. Minimally, it functions as an arbiter in the regular flow of these processes. We intervene when our conscious, active, self judges it fit to do so. Sometimes this executive self intervenes to resolve conflicts. Sometimes it intervenes to derail a mix of other mechanisms that have converged on a conclusion that, upon reflection, is consciously rejected. But conscious deliberation and the corresponding exercise of active agency do not always involve themselves solely to turn the present psychological tide. Sometimes we are forward-looking, setting up plans or weighing up values that structure downstream operation.

Much of the time it is obvious what the agent should do, and what way counts as a satisfactory way of doing it. Amongst adults it may frequently be the case that conscious deliberation only injects itself into the psychological tide when there is a special reason to do so. Such economy of intervention is oftentimes a good thing. Conscious deliberation is slow and demanding of neurochemical resources. Like all mechanisms, it is capable of error. Even so, to the extent to which it effectively resolves conflicts and sets in motion constraints on deliberation and action through planning and related mechanisms of psychological disciplining, it has an important role to play.

Situationism suggests that the empirical facts of our agency are at odds with our self-conception, that context matters more than we tend to suppose. The picture of free will I have offered is an attempt to be responsive to those facts. The resultant account is therefore likely to also be at some remove from our naïve self-conception. For example, the tendency to think that our capacities for control are metaphysically robust, unified, and cross-situationally stable is not preserved by my account of free will. Instead, free will involves capacities that are functions of agents, a context of action, and normatively structured practices. It is simply a mistake to think of free will as a kind of intrinsic power of agents.

Notice that this means free will is partly insulated from direct threats arising from experimental research. No matter how much we learn about the physical constitution of agency, it is exceedingly difficult to see how science alone could ever be in a position to settle whether some or another arrangement of our physical nature has normative significance for the integrity

of attributions of praise and blame.

9. From threat to tool

Current work in the psychological sciences threatens conventional Reasons accounts. I have argued that these threats can, to a large extent, either be shown to be less serious than they may appear or they can be met by abandoning some standard presumptions about what free agency requires (e.g., abandoning monism and atomism). I now wish to turn the argument on its head. Rather than thinking of the psychological sciences as presenting threats to free will, we do well to think of them as providing us with resources for enhancing what freedom we have.

There are at least two ways in which the data might guide us in the quest for enhanced control and the construction of circumstances conducive to our considerations-mongering agency. First, experimental results may tell us something about the conditions under which we are better and worse at perceiving moral considerations. Second, experimental results can illuminate the conditions under which we are better and worse at translating our commitments into action. Admittedly, translating the discoveries of experimental work into practical guidelines for moral improvement will always be a dodgy affair. Still, there are some simple reasons for optimism about the utility of these data.

One way of being better at perceiving moral considerations is to simply avoid things that degrade our those perceptual capacities. For example, suppose *Samaritan*-like data comes to show that for most people time pressure significantly reduces our capacity to perceive what we regard as moral considerations. Such a discovery has an important implication for our moral ecology: in those cases where we have an interest in not significantly limiting helping behavior, it behooves us to limit time-pressuring forces.

A related way data might prove to be useful is simply by making us aware of how situational effects influence us. There is some evidence that knowledge of situational effects can, at least sometimes, help reduce or eliminate their deleterious effects (Pietromonaco and Nisbett 1992; Beaman et al. 1978). These are simple, but suggestive ways in which data and philosophical theory might work together to limit our irrationality.

More ambitiously, there is intriguing evidence to the effect that situations do not merely contain the power to degrade our baseline capabilities,

but that they may *enhance* our capacity to detect at least some varieties of considerations. For example, when individuals perceive bias in their favor, it can actually *enhance* some cognitive tasks, and not just because of motivational effects.[27] For example, Aronson et. al. reported that in one study, males "performed worse in conditions where the female stereotype was nullified by experimental conditions. Specifically, males tended to perform worse when told that the test was not expected to show gender differences, suggesting that their performance may be boosted by the implicit stereotype" (42). In other words, stereotypes about male math advantage seem to benefit males in contexts where the stereotype is operating, as opposed to neutral contexts that counter stereotype bias. Thus far, there is disappointingly little experimental data on this phenomenon, which we might call *stereotype advantage.* Nevertheless, it points to another way in which the data, contrary to threatening us, might instead serve to contribute to the powers of our agency.

This possibility obviously raises troubling issues. It is surely morally problematic to exploit false stereotypes for cognitive advantage. Moreover, there are practical challenges to equitably exploiting stereotypes that are indexed to subsets of the larger population in a widespread way. Nevertheless, the possibility of non-problematic cognitive enhancers to our baseline capacities is worth further consideration.

Experimental data might also affect how we go about fostering responsible agency concerns the translation of principles into action. As we have seen, one of the disturbing elements in the Milgram studies, and more recently, in the Abu Ghraib tortures, is the suggestion that ordinary people can be led to behave abhorrently. The combination of the ordinariness of the perpetrators with the uncommonness of the atrocity is precisely what is so striking about these cases (Doris and Murphy 2007). One way of understanding such instances is that situational effects do not necessarily undo one's grasp of moral considerations (although they may do this, too), but that at least sometimes, they weaken the connection between conviction and action.

Issues here are complex. In the real world, distinguishing between

---

27 Claude Steele, et. al note doubt it is a motivation effect because "the effort people expend while exercising stereotype threat on a standardized test has been measured in several ways: how long people work on the test, how many problems they attempt, ho much effort they report putting in, and so on. But none of these has yielded evidence, in the sample studied, that stereotype threat reduces test effort" (Steele et al. 2002).

perceptual and motivational breakdowns may be difficult. Moreover, there are complex issues here concerning weakness of will and the point at which we do not expect people to satisfy the normative demands to which we ordinarily take them to be a subject.[28] For that matter, there is a lively discussion in psychology concerning the extent to which our conscious attitudes explain our behaviors (McGuire 1985; Kraus 1995). One thing that situationism strongly suggests, however, is that circumstances make a difference for the ability of agents to control their behaviors in light of their principles.[29] To the extent that responsible agency requires that an agent's attitudes control the agent's behaviors, experimental data can again provide some guidance on how we might better shape our environments to contribute to that control.

In sum, while situationist data might initially appear to threaten our freedom and moral responsibility, what we have seen is, if not quite the opposite, at least something considerably less threatening. Given a situation-sensitive theory of responsible agency and some attention to the data, we find that our agency is somewhat different than we imagine. The situationist threat turns out to be only one aspect of a more complex picture of the forces that enhance and degrade our agency. Whether and where we build bulwarks against the bad and granaries for the good is up to us.

Here, a suitably cynical critic might retort that this is indeed something, but not enough. After all, we are still faced with a not-altogether-inspiring upshot that since we do not control our situations as much as we like, we are still not responsible agents as much as we might have hoped.

At this point, a concession is in order. I agree that we have less freedom than we might have hoped for, but I must insist that we have more freedom than we might have feared. Although we must acknowledge that our

---

28 For my part, I do not think there is anything like a unified account to be told of the justified norms governing what counts as weakness of will and culpable failure. I suspect that these norms will vary by context and agent in complex ways, and in ways that are sensitive to folk expectations about psychological resilience and requirements on impulse management. In his characteristically insightful way, Gary Watson may have anticipated something like this point in the context of addiction: "the moral and legal significance of an individual's volitional weaknesses depends not only on judgments about individual responsibility and the limits of human endurance but on judgments about the meaning and value of those vulnerabilities" (Watson 2004, p. 347)

29 For example, in one study, the subjects' attitudes controlled their behavior more when he or she was looking at himself in a mirror (Carver 1975).

freedom-relevant capacities are jeopardized outside of responsibility-supporting circumstances, we may still console ourselves with the thought that we have a remarkable amount of control in suitable environments.

Such thoughts do not end the matter. The present equilibrium point gives rise to new issues worth mentioning, if only in closing. In particular, it is important to recognize that societies, states, and cultures all structure our actual capacities. Being raised in an anti-racist context plays a role in enhancing sensitivity to moral considerations tied to anti-racist concerns. Similarly, being raised in a sexist, fascist, or classist culture will ordinarily shape a person's incapacities to respond to egalitarian concerns. Such considerations may suggest that we need to ask whether societies or states have some kind of moral, practical, or political obligation to endeavor to shape the circumstances of actors in ways that insulate them against situational effects that degrade their (moral or other) reasoning. We might go on to ask whether societies or states have commensurate obligations to foster contexts that enhance our rational and moral agency. If they do, it suggest that free will is less a matter of science than it is of politics or morality.[30]

30 Thanks to Henrik Walter and Dana Nelkin (twice over) for providing both comments on predecessor papers as well as affording me the circumstances in which I couldn't put off writing about these things. Thanks also to Kristin Drake, Eddy Nahmias, Joshua Knobe, David Velleman, Till Vierkant, and David Widerker for helpful feedback on ideas in this paper. I am also grateful to Ruben Berrios and Christian Miller for their commentaries at the *Selfhood, Normativity, and Control* conference in Nijmegen and the Pacific APA in 2007, respectively; thanks, too, to audience members in both places.

References

Aronson, Joshua, Michael J. Lustina, Catherine Good, Kelli Keough, Claude M. Steele, and Joseph Brown. 1999. When White Men Can'T Do Math: Necessary and Sufficient Factors in Stereotype Threat. *Journal of Experimental Social Psychology* 35 29-46.

Arpaly, Nomy. 2003. *Unprincipled Virtue*. New York: Oxford.

Asch, Solomon. 1951. *Effects of Group Pressures Upon the Modification and Distortion of Judgment*. In *Groups, Leadership, and Men*, edited by Harold Guetzkow. Pittsburgh: Carnegie Press.

Bargh, John A. 2008. *Free Will is Un-Natural*. In *Are We Free? Psychology and Free Will*, edited by John Baer, James C. Kaufman, and Roy F. Baumeister. New York: Oxford University Press.

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.

Bayne, Tim. 2006. *Phenomenology and the Feeling of Doing: Wegner on the Conscious Will*. In *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*, edited by S. Pockett, W.P. Banks, and S. Gallagher. Cambridge, MA: MIT Press.

Beaman, A.L., P.L. Barnes, and B. McQuirk. 1978. Increasing Helping Rates Through Information Dissemination: Teaching Pays. *Personality and Social Psychology Bulletin* 4

Carver, C.S. 1975. Physical Aggression as a Function of Objective Self-Awareness and Attitudes Towards Punishment. *Journal of Experimental Social Psychology* 11 510-19.

Clarke, Randolph. 2009. Dispositions, Abilities to Act, and Free Will: The New Dispositionalism. *Mind* 118 (470): 323-51.

Darley, John, and Daniel Batson. 1973. 'From Jerusalem to Jericho': A Study of Situational and Dispositional Variables in Helping Behavior. *Journal of Personality and Social Psychology* 27 100-08.

Dennett, Daniel. 1984. *Elbow Room*. Cambridge: MIT.

Doris, John. 1998. Persons, Situations, and Virtue Ethics. *Nous* 32 504-30.

Doris, John. 2002. *Lack of Character*. New York: Cambridge University Press.

Doris, John, and Dominic Murphy. 2007. From My Lai to Abu Ghraib: The Moral Psychology of Atrocity. *Midwest Studies in Philosophy* 31 25-55.

Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Harman, Gilbert. 1999. Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society* 99 (3): 315-31.

Isen, Alice, and Paula Levin. 1972. Effect of Feeling Good on Helping. *Journal of Personality and Social Psychology* 21 384-88.

Kamtekar, Rachana. 2004. Situationism and Virtue Ethics on the Content of Our Character. *Ethics* 114 (3):

Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford.

Kihlstrom, John F. 2008. *The Automaticity Juggernaut—Or, Are We Automatons 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.

King, Matt. 2009. The Problem With Negligence. *Social Theory and Practice* 577-95.

Kraus, Stephen. 1995. Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature. *Personality and Social Psychology Bulletin* 21 58-75.

Latané, Bibb, and Judith Rodin. 1969. A Lady in Distress: Inhibiting Effects of Freinds and Strangers on Bystander Intervention. *Journal of Experimental Social Psychology* 5 189-202.

Li, Wen, Isabel Moallem, Ken A. Paller, and Jay A. Gottfried. 2007. Subliminal Smells Can Guide Social Preferences. *Psychological Science* 18 (12): 1044-49.

McGuire, W. J. 1985. *Attitudes and Attitude Change*. In *The Handbook of Social Psychology*, edited by G. Lindzey, and E. Aronson. New York: Random

House.

Mele, Alfred R. 2009. *Effective Intentions: The Power of the Conscious Will*. New York: Oxford University Press.

Merritt, Maria. 2000. Virtue Ethics and Situationist Personality Psychology. *Ethical Theory and Moral Practice* 3 365-83.

Milgram, Stanley. 1969. *Obedience to Authority*. New York: Harper and Row.

Miller, Christian. 2009a. Empathy, Social Psychology, and Gobal Helping Traits. *Philosophical Studies* 142 247-75.

Miller, Christian. 2009b. Social Psychology, Mood, and Helping: Mixed Results for Virtue Ethics. *Journal of Ethics* 13 145-73.

Moll, Jorge, Roland Zahn, R de Olliveira-Souza, Frank Krueger, and Jordan Grafman. 2005. The Neural Basis of Human Moral Cognition. *Nature Reviews Neuroscience* 6 799-809.

Montague, P. Read. 2008. Free Will. *Current Biology* 18 (14): R584-R585.

Nahmias, Eddy. 2002. When Consciousness Matters: A Critical Review of Daniel Wegner's the Illusion of Conscious Will. *Philosophical Psychology* 15 (4): 527-41.

Nahmias, Eddy. 2007. *Autonomous Agency and Social Psychology*. In *Cartographies of the Mind: Philosophy and Psychology in Intersection*, edited by Massimo Marraffa, Mario De Caro, and Francesco Ferretti. Berlin: Springer.

Nelkin, Dana. 2005. Freedom, Responsibility, and the Challenge of Situationism. *Midwest Studies in Philosophy* 29 (1): 181-206.

Nelkin, Dana. 2008. Responsibility and Rational Abilities: Defending and Asymmetrical View. *Pacific Philosophical Quarterly* 89 497-515.

Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Nisbett, Richard E. 2003. *The Geography of Thought: How Asians and Westerners Think Differently-- and Why*. New York: Free Press.

Pelham, Brett W., Matthew C Mirenberg, and John T. Jones. 2002. Why Susie Sells Seashells By the Seashore: Implicit Egotism and Major Life

Decisions. *Journal of Personality and Social Psychology* 82 (4): 469-87.

Perry, John. 2010. Wretched Subterfuge: A Defense of The Compatibilism of Freedom and Natural Causation. *Proceedings and Addresses of the American Philosophical Association* 84 (2): 93-113.

Pietromonaco, P, and Richard Nisbett. 1992. Swimming Upstream Against the Fundamental Attribution Error: Subjects' Weak Generalizations From the Darley and Batson Study. *Social Behavior and Personality* 10 1-4.

Sabini, John, Michael Siepmann, and Julia Stein. 2001. The Really Fundamental Attribution Error in Social Psychological Research. *Psychological Inquiry* 12 1-15.

Steele, Claude M, Steven J. Spencer, and Joshua Aronson. 2002. Contending With Group Image: The Psychology of Stereotype and Social Identity Threat. *Advances in Experimental Social Psychology* 34 379-440.

Talbert, Matthew. 2009. Situationism, Normative Competence, and Responsibility for Wartime Behavior. *Journal of Value Inquiry* 43 (3): 415-32.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin.

Vargas, Manuel. 2009. Reasons and Real Selves. *Ideas y Valores: Revista colombiana de filosofía* 58 (141): 67-84.

Vargas, Manuel. 2010. Responsibility in a World of Causes. *Philosophic Exchange* 40: 56-78.

Vargas, Manuel. 2011. *The Revisionist Turn: Reflection on the Recent History of Work on Free Will*. In *New Waves in the Philosophy of Action*, edited by Jesus Aguilar, Andrei Buckareff, and Keith Frankish. Palgrave Macmillan.

Vargas, Manuel, and Shaun Nichols. 2007. Psychopaths and Moral Knowledge. *Philosophy, Psychiatry, and Psychology* 14 (2): 157-62.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Watson, Gary. 2004. *Excusing Addiction*. In *Agency and Answerability*, New York: Oxford University Press.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Weyant, J. 1978. Effects of Mood States, Costs, and Benefits on Helping. *Journal of Personality and Social Psychology* 36 1169-76.

Wolf, Susan. 1990. *Freedom Within Reason*. New York: Oxford University Press.

Woolfolk, Robert L., John Doris, and John Darley. 2006. Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility. *Cognition* 100 283-401.