https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

SITUATIONISM, RESPONSIBILITY, AND FAIR OPPORTUNITY*

By David O. Brink

We tend to view ourselves and other normal, competent adult agents as having comparatively stable characters and as being responsible for our choices and their foreseeable consequences. We may not always behave consistently, if only because we experience occasional weakness of will, and our characters are not fixed but can and do change over time. Nonetheless, we do think of ourselves as having relatively settled character traits that shape much of our behavior. Moreover, we assume that acting from character is compatible with being responsible for our actions. As long as our character has not been deformed in ways that deprive us of competence to recognize and respond to reasons, we assume that we are responsible both when we act within character and when we succumb to weakness of the will. Competent agents are responsible for their conduct, except in unusual circumstances in which they act under extreme duress, for instance, as the result of coercion.

The situationist literature in psychology may seem to challenge these assumptions about character and responsibility. Situationism claims that conduct is not determined by character but reflects the operation of underappreciated aspects of the agent's situation or environment. For instance, several situationist studies suggest that our willingness to engage in helping behavior for those in need is determined to a very surprising extent by modest contextual factors that influence our mood. According to situationism, situational factors, as well as character, explain behavior. According to some situationist views, context, rather than character, is explanatorily primary. Situationist studies in which behavior seems better explained by contextual than by characterological factors can be both surprising and unsettling. They may suggest that we do not understand ourselves and our motivations very well, and they may raise questions about our competence and responsibility for our conduct.

Recently, several philosophers have embraced situationist findings and drawn revisionary philosophical conclusions, expressing skepticism about the existence of stable character traits, the prospects for virtue ethics, and

* I would like to thank Craig Agule, David Estlund, Erin Kelly, Brian Leiter, Michael McKenna, Dana Nelkin, Daniel Russell, and David Shoemaker for helpful comments and suggestions.

doi:10.1017/S026505251300006X
© 2013 Social Philosophy & Policy Foundation. Printed in the USA.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

ordinary assumptions about moral responsibility.[1] John Doris has developed these arguments about the revisionary implications of situationism for ethical theory and moral responsibility further than others.[2] In *Lack of Character* Doris summarizes situationist findings and argues that ordinary assumptions about character and its explanatory significance must be abandoned. He argues that this undermines any conception of virtue ethics that attaches moral significance to character and suggests that situationism challenges familiar conceptions of moral responsibility. In "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity" Doris and Dominic Murphy appeal to situationist ideas to motivate a wide-ranging excuse for wrongdoing during wartime, significantly limiting the scope of moral responsibility.[3]

I want to assess whether and, if so, in what ways situationism should revise our assumptions about moral responsibility. While I do think that situationism is relevant to moral responsibility and may lead to some revisions in our assumptions about responsibility at the margins, especially in wartime, I do not think that situationism's implications are anywhere near as revisionary as Doris does. Ethical theory can take situationism on board without capsizing the boat. It might be possible to explore the implications of situationism for responsibility without committing to any particular conception of responsibility, relying only on ecumenical assumptions about responsibility. By contrast, my approach is to see what situationism implies about responsibility conceived in a particular way that I think is plausible.[4] The approach I have in mind conceives of responsibility as requiring the fair opportunity to avoid wrongdoing, which is conceived as factoring into requirements of normative competence and situational control. Though not completely ecumenical, this approach is far from idiosyncratic and synthesizes familiar ideas from reasons-responsive approaches to moral responsibility and

[1] See John Doris, *Lack of Character* (New York: Cambridge University Press, 2002); Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99 (1999): 315–31; and Peter Vranas, "The Indeterminacy Paradox: Character Evaluations and Human Psychology," *Nous* 39 (2005): 1–42. For an earlier and less revisionary philosophical discussion of situationist findings, see Owen Flanagan, *Varieties of Moral Personality* (Cambridge: Harvard University Press, 1991).

[2] Doris has developed the arguments that interest me the furthest and drawn the most revisionary conclusions, and so I will focus on his discussion. Much of my discussion of Doris applies to Harman. Vranas's discussion is narrower and in important respects more circumspect and nuanced. My reservations about Doris's arguments may not apply to Vranas.

[3] John Doris and Dominic Murphy, "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity," *Midwest Studies in Philosophy* 31 (2007): 25–55.

[4] In this way, my response to Doris's situationist arguments differs from the response offered in Dana Nelkin, "Freedom, Responsibility, and the Challenge of Situationism," *Midwest Studies in Philosophy* 29 (2005): 181–206. Nelkin contrasts two ways of responding to the situationist challenge to responsibility—one is to measure the significance of situationist findings against a particular conception of responsibility, while another is to ask, prior to the adoption of a particular conception of responsibility, why situationist findings might seem to threaten responsibility and then to see if those worries are well founded. Whereas Nelkin pursues the second strategy, I pursue the first.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

fair choice approaches to criminal responsibility. It is easier, I think, to reach determinate conclusions about the bearing of situationism on responsibility if we operate with a reasonably determinate conception of responsibility.

Though my main focus is on responsibility, some attention must also be paid to other forms of ethical significance that situationism is alleged to have. Understanding the limits of situationism's reach on some other ethical issues will help us see the limitations of its reach on issues about responsibility.

My plan is to begin by sketching some pertinent situationist ideas and findings. Then I turn to the broader ethical significance of situationism, suggesting that Doris's more revisionary conclusions can and should be resisted. To assess the bearing of situationism on responsibility, we need assumptions about responsibility. I sketch an attractive framework for responsibility that understands it as requiring the fair opportunity to avoid wrongdoing, where that in turn requires both normative competence and situational control. Against this background, I argue that in paradigmatic contexts studied by situationists there is little reason to deny that agents are responsible for their conduct. Finally, I discuss the bearing of situationism on wartime wrongdoing. Here, I agree that the heat of battle can create situational pressures that significantly impair normative competence and thus provide a full or partial excuse. But I express skepticism about Doris's and Murphy's tendency to generalize this kind of excuse to other contexts of wartime wrongdoing. As this outline suggests, my discussion ranges widely and must be impressionistic and selective in places. But the issues are genuinely interdependent and need to be discussed together. I hope that the value of treating them together and the plausibility of the general line of argument compensate for the need to be impressionistic and selective.

I. Situationism

The situationist literature contains a variety of experiments that are typically held to show that personality traits explain far less behavior, and that situational factors, sometimes apparently inconsequential ones, explain far more behavior, than is often supposed. This has led some psychologists to identify the so-called "Fundamental Attribution Error," which Ross and Nisbett characterize as involving:

> [p]eople's inflated belief in the importance of personality traits and dispositions, together with their failure to recognize the importance of situational factors in affecting behavior.[5]

[5] Lee Ross and Richard Nisbett, *The Person and the Situation* (Philadelphia: Temple University Press, 1991), 4.

A series of situationist studies support the Fundamental Attribution Error. Many of these studies share general features. They examine character traits, such as compassion, that are assumed to be fairly common and widespread. They place subjects in situations that would seem likely to elicit the corresponding behavior, and they study ways in which apparently modest situational or contextual factors affect the behavior of subjects, inducing them to behave in unexpected ways. Here are a few landmark studies.

1. *Group Effects.* Solomon Asch performed experiments to test whether the tacit desire for conformity with one's peers would induce subjects to report obviously wrong answers to perceptual questions.[6] In the Asch paradigm a real subject was placed with confederates and they were asked questions about the comparative lengths of various lines. The confederates answered first and gave the same answers as each other. After giving a few correct answers, the confederates began giving obviously wrong answers. In a control group, in which there was no pressure to conform, only one subject in 35 ever gave a wrong answer. However, in cases in which confederates gave the wrong answer first, subjects gave incorrect answers 32 percent of the time. In fact, 75 percent of subjects gave an incorrect answer to at least one question in these circumstances.

2. *Compassionate Behavior.* Several studies have shown compassionate behavior to be correlated with modest situational factors. (A) Alice Isen and Paula Levin performed an experiment in which they tested the helping behavior of subjects who had just made a telephone call in a public phone booth.[7] Some subjects found a dime that had been planted in the change slot by the experimenters; for other subjects, no dime was left. All subjects encountered a confederate while leaving the phone booth who dropped a folder full of papers, creating an opportunity for the subject to provide assistance. Only one in 24 subjects who did not find a dime helped; 16 of 18 who did find a dime helped. (B) Jon Darley and Daniel Batson performed an experiment involving seminary students who had been invited to prepare and deliver a talk on either the parable of the Good Samaritan or possible employment for seminary students.[8] En route to the talk, the seminary students encountered a confederate slumped in a doorway and in apparent need of medical attention. Despite the subject of their talks, the best predictor of whether subjects would stop and help was whether they had been told they should hurry to get to their talk. Subjects told they were running ahead engaged in helping behavior

[6] Solomon Asch, "Effects of Group Pressures on the Modification and Distortion of Judgment," in *Groups, Leadership, and Men,* ed. H. Guetzkow (Pittsburgh: Carnegie Press, 1951).

[7] Alice Isen and Paula Levin, "Effect of Feeling Good on Helping," *Journal of Personality and Social Psychology* 21 (1972): 384–88.

[8] Jon Darley and Daniel Baton, "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27 (1973): 100–108.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

63 percent of the time; subjects told they were on time engaged in helping behavior 45 percent of the time; and subjects told they were running late helped only 10 percent of the time.

3. *Authority.* Among the most famous of situationist experiments is a series of experiments performed by Stanley Milgram in which subjects engage in apparently harmful behavior in response to noncoercive instructions from the experimenter.[9] Subjects were volunteers in an experiment ostensibly about memory and learning. An experimenter asked each subject to participate in an experiment with another person who was in fact a confederate of the experimenter. The confederate was to be the learner and was asked questions by the subject. In the basic format of the experiment, the learner was placed in another room, apparently hooked up via electrodes to a device that the subject was told would deliver shocks that the subject would be asked to administer in response to the learner's mistakes. The shocks were staged, not real. After each mistake, the subject was asked to administer a shock to the learner, increasing the magnitude of the shock after each mistake. The shocks were represented to the subject as ranging from slight (15–60 volts), to moderate (75–120 volts), to strong (135–80 volts), to very strong (195–240 volts), to intense (255–300 volts), to extremely intense (315–60 volts), and to dangerous (375–420 volts). In the standard format of the experiment, as the apparent voltage was increased, the confederate responded with increasing (feigned) distress and finally (and ominously) with silence. If subjects expressed concern about continuing to administer apparently more serious shocks, the experimenter responded by politely but firmly insisting that the experiment was important and that the shocks were painful but not harmful. In the original study, 62.5 percent of subjects shocked their learners to the maximum allowable amount into the dangerous zone with intense protestation and screams from the confederate followed by silence. These results held up in numerous replications and permutations of the original experiment.

4. *Role Playing*. Equally infamous was Philip Zimbardo's Stanford Prison Experiment in which a simulated prison was set up in the basement of the Stanford Psychology Department and in which student volunteers were randomly assigned roles as prisoners and guards.[10] Though they were asked to play the roles of prisoners and guards, guards were instructed to avoid physical punishment and to treat their charges humanely. The experiment was to run for two weeks but had to be terminated after six days because the guards had become increasingly abusive and sadistic and prisoners were experiencing genuine psychological trauma.

In all these studies, people displayed surprising and unsettling patterns of behavior, and there is reason to think that what explains these surpris-

[9] Stanley Milgram, *Obedience to Authority* (New York: Harper & Row, 1969).

[10] The Stanford Prison Experiment took place in the early 1970s. See Philip Zimbardo, *The Lucifer Effect* (New York: Random House, 2007).

ing results are situational factors, factors that we might not have thought would be so important. This is why situationists identify a Fundamental Attribution Error. The Fundamental Attribution Error is a *comparative* claim. In fact, it does not even make the comparative claim that situation is a bigger behavioral determinant than character. It claims only that most people tend to overestimate the importance of character and underestimate the importance of situation. This might be true even if character is more explanatory, provided that it is less explanatory than many assume.

II. Situationism, Character, and Virtue

In *Lack of Character* Doris goes much further than the Fundamental Attribution Error, not just claiming that situation is more important and character less important than we thought, but also expressing skepticism about the existence of stable character traits.[11] We might distinguish three possible situationist claims of increasing strength.

(1) Character is less important and situational factors are more important than we thought in explaining human behavior.
(2) Character is less important than situational factors in explaining human behavior.
(3) There are no stable character traits.

The Fundamental Attribution Error makes the interesting but comparatively weak claim in (1). A stronger comparative claim would be (2). Neither of these comparative claims implies skepticism about character traits of the sort expressed in (3) and endorsed, at least sometimes, by Doris.

Doris's skepticism about character reflects three assumptions that he thinks are pervasive in characterological moral psychology.

(1) *Cross-situational consistency.* Character traits are cross-situationally (or horizontally) consistent behavioral dispositions. If a person has a trait, it should manifest itself consistently in a variety of behavioral contexts.
(2) *Diachronic Stability.* This is a temporal (vertical, rather than horizontal) behavioral consistency requirement, which is not cross-situational.
(3) *Evaluative Integration.* If a person has one morally relevant character trait, she is more likely to have others. The inseparability of the virtues, according to which if a person has one virtue she must have all the others, would be an extreme version of evaluative integration.

[11] I provide references to Doris, *Lack of Character* parenthetically in the text hereafter.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

Doris concludes that situationism undermines cross-situational consistency and evaluative integration but leaves diachronic stability intact (pp. 22–23). Most agents do not display cross-situationally consistent character traits, and they tend to be morally fragmented. Only diachronically consistent local traits exist (pp. 62–67).

Doris goes on to develop two different implications of situationism for ethical theory, one negative and one positive. The negative ethical conclusion is that virtue ethics and other forms of moral character evaluation are bankrupt because they rest on mistaken assumptions about the existence of robust character traits (pp. 107–19). The positive thesis is that it might be possible to remake a characterological moral theory that is sensitive to the findings of situationism (an empirically adequate characterological theory) or, if that is not possible, to fashion a non-characterological moral theory that helps facilitate better moral outcomes by attending to the importance of situational factors in behavior (pp. 120–21, 146, 163).

Unfortunately, Doris focuses primarily on the negative thesis and does not develop the constructive lessons of situationism in much detail. This is a missed opportunity. We might want to modify our institutions, policies, and personal plans in light of a recognition of the importance of situational factors. For instance, the Fundamental Attribution Error might help me recognize that I am less likely to succeed in dealing with substance abuse problems by sheer will power than by removing the availability of the substance (e.g., throwing the liquor bottles out) and changing my patterns of association (e.g., no longer socializing with my drinking buddies in contexts in which alcohol is served). Situationism might also provide further support for ideas in behavioral economics about how the context or framework of choice can influence outcomes. For instance, Cass Sunstein and Richard Thaler advocate a regulatory approach that they call "libertarian paternalism," because it structures choice in ways that tend to lead to better welfare outcomes for those making choices but in ways that do not limit choice.[12] We could *nudge,* rather than coerce. We could arrange for the healthy foods in the cafeteria to be more easily accessible, reserving the less accessible locations for less healthy foods. Sometimes, we can nudge just by changing the default options in a menu. In the case of defined contribution retirement plans, we could make the maximum annual contribution the default option, which the agent could change but only through deliberate choice. The way we manage the framework of choice should take into account situational influences on choice. The missed opportunity of exploring the constructive implications of situationism will be all the more important if, as I believe, Doris overstates his negative case.

[12] See Cass Sunstein and Richard Thaler, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New Haven: Yale University Press, 2008).

III. The Ethical Implications of Situationism

Doris's negative argument raises a number of issues. He infers skepticism about character from inconsistent behavioral manifestation. One might well question whether behavioral inconsistency undermines the ascription of character traits and whether familiar moral assumptions presuppose cross-situationally stable traits.

It is not clear that we should draw skeptical conclusions about character from data about behavioral inconsistency. In fact, a significant strand in the social psychology literature expressly resists this skeptical conclusion, treating situationist findings as motivating a situationally adequate and constrained conception of character, one in which character is not a simple behavioral disposition but a complex suite of dispositions to perceive and attend to certain aspects of one's situation, to feel certain emotions, and to react appropriately.[13]

Despite some disclaimers (pp. 15–16), Doris focuses almost exclusively on behavior and treats failures to manifest the trait in certain conditions as evidence that the person does not have a robust trait. But traits might manifest in other ways, for instance, in emotional responses, such as anxiety or regret, or in compensating behavior, such as apology, restitution, and learning from mistakes.[14] For instance, it is noteworthy that most of the compliant subjects in Milgram experiments manifested visible signs of stress and anxiety as they complied with requests to administer shocks. These other forms of manifestation might serve as evidence of a character trait, even in the absence of invariant behavioral manifestations. We might wish that the character trait was stronger and manifested itself in compassionate behavior more reliably, but that does not show that the subject does not possess the trait.

Furthermore, traits might display a behavioral tendency and yet be *masked* by other traits. If traits are, in part, dispositions that tend to produce behavior when other things are equal and there are no interfering conditions, then we should not expect traits to manifest themselves behaviorally in situations in which there is significant situational interference. A given situation might engage more than one trait. If the traits tend

[13] See Walter Mischel, *Personality and Assessment* (Hoboken: John Wiley & Sons, 1968) and Ross and Nisbett, *The Person and the Situation*. For a philosophical analysis of this strand in the empirical work, see Daniel Russell, *Practical Intelligence and the Virtues* (Oxford: Clarendon Press, 2009), esp. chaps. 8–9.

[14] Several philosophers have raised questions about Doris's behavioral assumptions about character traits and his demand for cross-situational consistency. See, for example, Robert Adams, *A Theory of Virtue* (Oxford: Clarendon Press, 2006), esp. chap. 8, and Russell, *Practical Intelligence and the Virtues*, esp. chaps. 8–9. One would like to know whether some of Milgram's subjects thought about their choices in the experiment and behaved differently, going forward, as the result of the experiment and the debriefing afterward. There is no reason to expect this sort of effect in unconflicted subjects, but it would be a real possibility for conflicted subjects. Milgram reports that some subjects later reported having learned important things about themselves from the experiment. See *Obedience to Authority*, 200.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

toward incompatible courses of conduct, then at least one trait will not manifest itself behaviorally. For instance, in the Milgram experiments, should we conclude from failures of compassionate behavior, as Doris does, that those who fail to display compassion are not compassionate, or should we conclude instead that people who are compassionate are also deferential to authority? Milgram interpreted his own experiments in the second way, as showing that deference is a real and powerful character trait. We might wish that the compassion of many subjects had been stronger and won out over their deference to authority, but the fact that they deferred does not itself mean that they are not compassionate. Similarly, the seminarian experiment need not show that the seminarians were not compassionate; instead, it shows that they recognized a special obligation to deliver their promised talks, perhaps assuming that others without special commitments could provide any needed assistance. Perhaps it is a condition of our ascribing a trait such as compassion to a person that it manifest itself behaviorally often enough. But how high that behavioral threshold should be is not clear, and it does seem clear that some failures to manifest the trait are compatible with possessing it.

We might also ask whether our normal assumptions about character and our ordinary moral discriminations are jeopardized by recognition that few people have cross-situationally stable character traits.

First, situationist experiments do not show that no one has cross-situationally consistent traits, just that they aren't as common as we might have thought or hoped. We could hold onto the idea that traits must be manifested behaviorally and simply conclude that virtue is less common than some think. Although nearly two-thirds of Milgram's subjects delivered the maximum shock, a little more than one-third did not. This is only so much comfort, because some of these subjects supplied intense shocks short of the maximal shock. But 20 percent did refuse to administer strong shocks, and 15 percent refused to continue as soon as the learner protested and asked for the experiment to stop. Did anyone think real virtue was more common than that?[15]

Second, situationism does not show that there are no differences in the moral behavior of people. Even if few people are as consistent as we might have thought or hoped, can't *comparative* differences between people still ground moral discriminations among them and their degree of virtue? From the fact that few are consistently or perfectly virtuous, it does not follow that the rest of us are all equally virtuous or nonvirtuous. If consistency and virtue are scalar, then we can still draw familiar moral distinctions among people based on their character, and even the frequency with which they manifest the trait in behavior, even if few have perfect character or virtue.

[15] Cf. Rachana Kamtekar, "Situationism and Virtue Ethics on the Content of our Character," *Ethics* 114 (2004): 458–91.

Third, knowledge is power. Once people recognize ways in which they are prone to situational influences, they may become better at resisting them. Indeed, it is precisely here that the positive potential of recognizing the role of situational factors can be empowering. In this connection, Milgram notes that some subjects later reported having become more aware of their susceptibility to influence by authorities and better able to resist authority when it conflicts with conscience as the result of the experiment.[16]

These responses raise questions about whether situationist findings really do undermine our normal assumptions about character and its moral importance. They suggest that Doris's skepticism is premature.

IV. Responsibility As the Fair Opportunity to Avoid Wrongdoing

To assess the implications of situationism for responsibility we have to have some assumptions about responsibility in place. So I would like to sketch a conception of the architecture of responsibility that draws on ideas in the reasons-responsive wing of the compatibilist tradition of thinking about free will and responsibility, and ideas from the fair choice approach to criminal responsibility.[17] This sketch will require us to gloss over and simplify some complex and potentially contested matters, with the result that our conclusions will be hostage to an incomplete conception of responsibility. But anything more would be beyond the scope of this essay, and I take some comfort in the fact that the conception on which I am relying is reasonably ecumenical, even if it is not uncontroversial in all quarters.

On a broadly retributive view, blame and sometimes punishment are fitting responses to wrongdoing for which the agent is culpable or responsible. More specifically, blame and punishment ought to be proportional to desert, which is itself the product of two independent variables—

[16] See Milgram, *Obedience to Authority*, 200.

[17] The reasons-responsive tradition of moral responsibility on which I draw is reflected in Susan Wolf, *Freedom within Reason* (New York: Oxford University Press, 1990); R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1994); John Fischer and Mark Ravizza, *Responsibility and Control* (New York: Cambridge University Press, 1998); and Dana Nelkin, *Making Sense of Freedom and Responsibility* (Oxford: Clarendon Press, 2011). The fair choice literature in criminal jurisprudence on which I draw is reflected in H. L. A. Hart, *Punishment and Responsibility* (Oxford: Clarendon Press, 1968); Michael Moore, *Placing Blame* (Oxford: Clarendon Press, 1997); and Stephen Morse, "Culpability and Control," *University of Pennsylvania Law Review* 142 (1994): 1587–1660 and "Uncontrollable Urges and Irrational People," *Virginia Law Review* 88 (2002): 1025–78. A fuller presentation of the conception of responsibility as the fair opportunity to avoid wrongdoing is contained in David O. Brink and Dana K. Nelkin, "Fairness and the Architecture of Responsibility," *Oxford Studies in Agency and Responsibility* 1 (2013): 284–313.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

wrongdoing and culpability or responsibility.[18] The view can be formalized as follows:

$$P \propto D\ (= W \times R)$$

The retributivist formula should be understood as claiming only that culpable wrongdoing is the desert basis for a *pro tanto* case for blame and punishment. That is, a plausible form of retributivism should allow that its desert-based rationale for blame and punishment can sometimes compete with and sometimes lose out to non-desert-based rationales against blame and punishment, appealing to factors such as mercy or forgiveness.[19]

This sort of broadly retributive view explains well the two main kinds of defense an agent might offer when blame or punishment threatens—*justifications* and *excuses.* Justifications deny wrongdoing, whereas excuses deny culpability or responsibility. Just as justification is the flipside of wrongdoing, so too excuse is the flipside of responsibility. These defenses are part of the criminal law but also reflect well the moral landscape. When others threaten to blame us, it can be appropriate to respond by citing factors that would justify or excuse our behavior. Our focus here is on responsibility and excuse. Excuses provide a window onto responsibility, and our conceptions of excuse and responsibility should have corresponding structure.

If someone is to be culpable or responsible for her wrongdoing, then she must be a responsible agent. Our paradigms of responsible agents are normal mature adults with certain sorts of capacities. Responsible agents must be *normatively competent.* They must be able to distinguish between the *intensity* and *authority* of their desires and impulses. This requires that agents not simply act on their strongest desires, but be capable of stepping back from their desires, evaluating them, and acting for good reasons. This requires responsible agents to be able to recognize and act on reasons for action. If so, normative competence involves *reasons-responsiveness,* which itself involves both *cognitive* capacities to distinguish right from wrong and *volitional* capacities to conform one's conduct to that normative knowledge.

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

[18] Here, I draw on related ideas in Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), 363–66.

[19] This is one of several ways in which forward-looking considerations can play a role in a conception of punishment that is constrained by backward-looking retributive concerns. Elsewhere, I describe this broadly retributive commitment as *predominant retributivism.* See, for example, David O. Brink, "Retributivism and Legal Moralism," *Ratio Juris* 25 (2012): 496–512.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

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. If responsibility were predicated on the proper use of these capacities, we could not hold weak-willed and willful wrongdoers responsible for their wrongdoing. Indeed, the fact of wrongdoing would itself be exculpatory, with the absurd result that we could never hold anyone responsible for wrongdoing. It is a condition of our holding wrongdoers responsible that they possessed the relevant capacities at the time of wrongdoing.

Normative competence, on this conception, involves two forms of reasons-responsiveness: an ability to recognize wrongdoing and an ability to conform one's will to this normative understanding. Both dimensions of normative competence involve norm-responsiveness. As a first approximation, we can distinguish moral and criminal responsibility at least in part based on the kinds of norms to which agents must be responsive.[20] Moral responsibility requires capacities to recognize and conform to moral norms, including norms of moral wrongdoing, whereas criminal responsibility requires capacities to recognize and conform to norms of the criminal law, including norms of criminal wrongdoing.

Normative competence requires the cognitive capacity to make suitable normative discriminations, in particular, to recognize wrongdoing. If responsibility requires normative competence, and normative competence requires this cognitive capacity, then we can readily understand one aspect of the criminal law insanity defense. A full account of the elements of insanity is controversial. But most plausible versions of the insanity defense include a cognitive dimension, first articulated in the M'Naghten rule that excuses if the agent lacked the capacity to discriminate right from wrong at the time of action.[21]

But there is more to normative competence than this cognitive capacity. It is a common view that intentional action is the product of informational states, such as beliefs, and motivational states, such as desires and intentions. Although our beliefs about what is best can influence our desires, producing optimizing desires, our desires are not always optimizing. Sometimes they are good-dependent but not optimizing, when they are directed at lesser goods, and sometimes they are completely good-independent. This is reflected in cases of weakness of will in which we have beliefs about what is best (and perhaps optimizing desires) but in which we act instead on the basis of independent non-optimizing passions and desires. This psychological picture suggests that being a responsible agent is not merely having the capacity to tell right from wrong but also requires the capacity to regulate one's actions in accordance with this

[20] I leave to another occasion the question whether there are other significant differences between moral and criminal responsibility.

[21] M'Naghten's Case, 10 Cl. & F. 200, 8 Eng. Rep. 718 (1843).

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

normative knowledge. This kind of volitional capacity requires emotional and appetitive capacities to enable one to form intentions based on one's optimizing judgments and execute these intentions over time, despite distraction, temptation, and other forms of interference.

If one's emotions and appetites are sufficiently disordered and outside one's control, this might compromise volitional capacities necessary for normative competence. Consider the following obstacles to volitional competence.

- Irresistible desires or paralyzing fears that are neither conquerable nor circumventable, as perhaps in some cases of genuine agoraphobia or addiction.[22]
- Clinical depression that produces systematic weakness of will in the form of listlessness or apathy.
- Acquired or late onset damage to the prefrontal cortex of the brain in which agents have considerable difficulty conforming to their own judgments about what they ought to do, as in the famous case of Phineas Gage.[23]

Each of these cases involves significant volitional impairment in which agents experience considerable difficulty implementing or conforming to the normative judgments they form.

Notice that recognition of a volitional dimension of normative competence argues against purely cognitive conceptions of insanity, such as the M'Naghten test, which recognizes only cognitive deficits as the basis for insanity, and in favor of the more inclusive *Model Penal Code* conception.

> Mental Disease or Defect Excluding Responsibility: (1) A person is not responsible for criminal conduct if at the time of such conduct as the result of a mental disease or defect he lacks substantial capacity either to appreciate the criminality [wrongfulness] of his conduct or *to conform his conduct to the requirements of law.* (2) [T]he terms "mental

[22] Cf. Al Mele, "Irresistible Desires," *Nous* 24 (1990): 455–72. Mele understands a desire as conquerable when one can resist it and as circumventable when one can perform an action that makes acting on the desire impossible or at least more difficult. The alcoholic who simply resists cravings conquers his impulses, whereas the alcoholic who throws out his liquor and stops associating with former drinking partners or won't meet them at places where alcohol is served circumvents his impulses. Conquerability is mostly a matter of will power, whereas circumventability is mostly a matter of foresight and strategy.

[23] Phineas Gage was a nineteenth-century railway worker who was laying tracks in Vermont and accidentally used his tamping iron to tamp down a live explosive charge, which detonated and shot the iron bar up and through his skull. Though he did not lose consciousness, over time his character was altered. Whereas he had been described as someone possessing an "iron will" before the accident, afterward he had considerable difficulty conforming his behavior to his own judgments about what he ought to do. The story of Phineas Gage is related, and its larger significance explored, in Antonio Damasio, *Descartes' Error: Emotion, Reasons, and the Human Brain* (New York: Putnam, 1994).

> disease or defect" do not include an abnormality manifested only by repeated criminal or otherwise antisocial conduct.[24]

The *Model Penal Code* conception of insanity is an important advance on the M'Naghten conception, precisely because it recognizes an independent volitional dimension to sanity and so recognizes a wider conception of insanity as involving significant impairment of *either* cognitive or volitional competence.

Both cognitive and volitional competence involve *sensitivity* to reasons. But sensitivity is a *scalar* notion. People can be more or less sensitive, and cognitive or volitional competence can be more or less impaired. If we analyze responsibility and excuse in terms of normative competence, this implies that both responsibility and excuse are, at least in principle, scalar concepts. There may be important pragmatic reasons for making bright line discriminations in certain contexts, privileging certain degrees of normative competence, as necessary and sufficient for responsibility and excuse. These are interesting and complicated issues, beyond the scope of this essay. But if we must recognize thresholds of normative competence, if only for pragmatic reasons, it seems fair to require substantial, rather than bare, normative capacity as a condition of responsibility and significant, rather than complete, impairment of capacity as grounds for an excuse. Hopefully, this kind of rough proportionality will be adequate for our purposes here.

An important part of an agent's being responsible for wrongdoing that she chose and intended consists in her being a responsible agent. We have conceptualized this in terms of normative competence and analyzed it into cognitive and volitional capacities. Evidence for this view is that one seems to have an excuse, whether complete or partial, if one's normative competence is compromised in significant ways. Some of the most familiar kinds of potential excuse—insanity, immaturity, and uncontrollable urges—involve compromised normative competence.

But there is more to an agent being culpable or responsible for her wrongdoing than her being responsible and having intentionally engaged in wrongdoing. Excuse is not exhausted by denials of normative competence. Among the factors that may interfere with our reactive attitudes, including blame and punishment, are *external* or *situational* factors. In particular, *coercion* and *duress* may lead the agent into wrongdoing in a way that nonetheless provides an excuse, whether full or partial. The paradigm situational excuse is coercion by another agent, as when one is threatened with physical harm to oneself or a loved one if one does not

[24] American Law Institute, *Model Penal Code,* section 4.01, emphasis added. The *Model Penal Code* (MPC) is a model statutory text of fundamental provisions of the criminal law, first developed by the American Law Institute in 1962 and subsequently updated in 1981. The MPC was intended to serve as a model for local jurisdictions drafting and revising their criminal codes.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

participate in some kind of wrongdoing, for instance, driving the getaway car in a robbery. The *Model Penal Code* adopts a reasonable-person version of the conditions under which a threat excuses, namely, when a person of reasonable firmness would have been unable to resist, provided the actor was not himself responsible for being subject to duress (section 2.09). Though criminal law doctrine focuses on threats that come from another's agency, hard choice posed by natural forces seems similarly exculpatory, as in Aristotle's famous example of the captain of the ship who must jettison valuable cargo in dangerous seas caused by an unexpected storm.[25] Situational duress does not compromise the wrongdoer's status as a responsible agent and does not challenge her normative competence, but it does challenge whether she is responsible for her wrongdoing.[26]

So far, the conception of responsibility emerging here is a two-factor model twice over. Responsibility is factored into normative competence and situational control, and normative competence is factored into cognitive and volitional capacities. This kind of two-factor model seems plausible, in significant part because it promises to fit our practices of excuse in both moral assessment and the criminal law pretty well. Perhaps this is adequate justification. But it would be nice if there were some unifying element to its structure.

One possible umbrella concept is *control.* Freedom from coercion and duress, cognitive competence, and volitional competence all seem to be aspects of an agent's ability to control her actions. But control seems important, at least in part, because it seems *unfair* to blame agents for outcomes that are outside their control. This suggests that the umbrella concept should be *fairness,* in particular, the fair opportunity to avoid wrongdoing, because failure of either normative competence or situational control violates the norm that blame and punishment be reserved for those who had a fair opportunity to avoid wrongdoing. If we treat the fair opportunity to avoid wrongdoing as the key to responsibility, we get the following picture of the architecture of responsibility (see Figure 1 on the next page).

[25] Aristotle, *Nicomachean Ethics,* trans. T. Irwin (Indianapolis: Hackett, 1985), 1110a9-12.

[26] The details of duress are tricky. Some situational pressures, such as the need to choose the lesser of two evils, may actually *justify* the agent's conduct, as is recognized in *necessity* defenses. If the balance of evils is such that the evil threatened to the agent is worse than the evil involved in her wrongdoing, then compliance with the threat is justified. But in an important range of cases, coercion and duress seem not to justify conduct (remove the wrongdoing) but rather to *excuse* wrongdoing, in whole or in part. In such cases, where the evil threatened is substantial but less than that contained in the wrongdoing, the agent's wrongdoing should be excused because the threat or pressure was more than a person could or should be expected to resist. Whether the difference between when duress justifies and when it excuses should be settled by applying the balance of evils test depends on the moral framework in which we measure evils. The criminal law tends to assume a consequentialist version of the test, but one might instead measure the moral seriousness of evils in ways that reflect agent-centered prerogatives.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

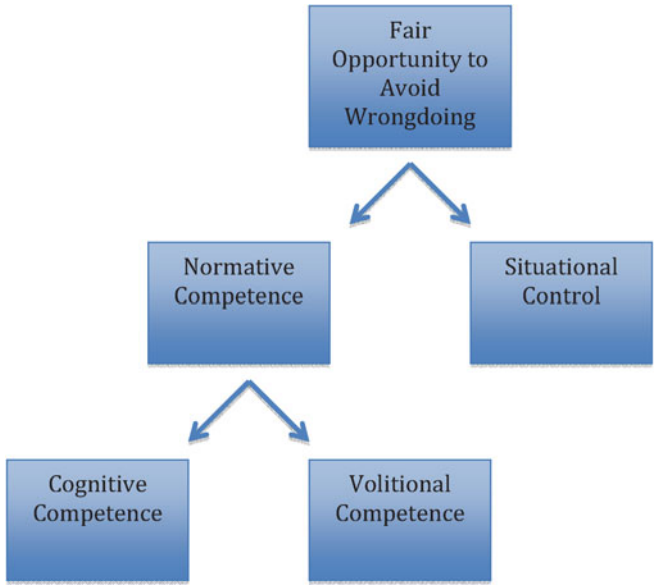


FIGURE 1. Fair Opportunity and the Architecture of Responsibility

One way to see the importance of the fair opportunity to avoid wrongdoing is to think about why strict liability is problematic. Strict liability offenses are those for which wrongdoing is sufficient and culpability is not required. It may be acceptable to impose liability without culpability in civil (tort) law. But in those parts of law and morals where we are interested in *accountability*, the idea of liability without culpability is problematic. In particular, broadly retributivist practices that condition blame and punishment on culpable wrongdoing must object to strict liability as unfair.[27]

[27] It is common to observe that strict liability is anomalous in the criminal law. Statutory rape laws that do not admit reasonable belief that the minor was an adult as a defense are sometimes cited as the main exception to the generalization that strict liability crimes are not recognized. Moreover, this particular conception of statutory rape is itself anomalous, recognized only in a minority of jurisdictions, and is widely criticized by commentators as unfair. However, someone might point to strict liability conceptions of traffic laws as further evidence of the willingness to tolerate strict liability crimes. To see why none of these laws is in fact a strict liability crime in the relevant sense, we must see that there are two distinct conceptions of strict liability crimes corresponding to two distinct conceptions of culpability. The criminal law conceives of wrongdoing as having two parts: *actus reus* and *mens rea.* The *actus reus* refers to the objective elements (action type, consequences, and attendant circumstances) that constitute the offense, whereas the *mens rea* refers to the mental element ingredient in the offense, such as whether the harmful act was intended or merely foreseen or was performed negligently or recklessly. Within the criminal law, culpability has two senses: in its narrow sense, culpability refers the mental elements that are ingredient in a crime (its *mens rea*); in its broader sense, culpability refers to blameworthiness or responsibility. Because culpability has these two distinct senses, and because strict liability crimes would be crimes in which there was liability without culpability, there are, in principle, two

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

V. Situationism and Responsibility

In the penultimate chapter of his book, Doris explores the implications of situationism for our assumptions about responsibility. His claim is that situationism poses problems for various conceptions of responsibility, especially *characterological* and *normative competence* conceptions. However, Doris seems to think that if we adopt his favored *identificationist* conception of responsibility, the problems situationism poses are manageable.[28]

A characterological conception of responsibility has some prima facie plausibility.[29] When someone acts in character, her actions express her settled beliefs and desires, and so seem to be attributable to her. By contrast, if an agent acts out of character, her actions do not express her true self, but perhaps the pressures of the situation or the influence of other people. So we might conclude that an agent is responsible for her action just in case it expresses her character or true self. One of the virtues of such a character theory is that it seems fully compatible with determinism, inasmuch as it makes responsibility a matter of a causal relation between actions and character. Responsible action not only can be but also *must be* caused by character, and character can itself be causally determined.

A characterological conception of responsibility will be problematic if it is committed to cross-situational consistency and if situationism shows that there are no such robust traits. If situationism supports only the

possible conceptions of strict liability crimes. (1) One kind of strict liability crime would be a crime that does not require any specific mental element as an ingredient in the offense. (2) Another kind of strict liability crime would be one that does not require blameworthiness or responsibility as a condition of blame and punishment. The wisdom of recognizing the first kind of strict liability crime is an interesting question, but there can be and are strict liability crimes in this sense. However, there are no strict liability crimes in the second sense, because insanity and duress, which are excuses denying responsibility (in different ways), are perfectly general defenses, applicable to all crimes. So even crimes that would be strict liability crimes in the first sense, requiring no particular *mens rea,* admit of defenses that establish insanity or duress. So while both kinds of strict liability crimes are problematic in the criminal law, the criminal law refuses to recognize the second kind of strict liability crime, because such crimes would be unfair.

[28] Interestingly, in "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity" Doris and Murphy rely exclusively on a normative competence conception of responsibility.

[29] Cf. David Hume's claims in *An Enquiry Concerning Human Understanding,* ed. P. H. Nidditch (1748; Oxford: Clarendon Press, 1975).

> The only proper object of hatred or vengeance is a person or creature, endowed with thought or consciousness; and when any criminal or injurious actions excite that passion, it is only by their relation to the person, or connection with him. Actions are by their very nature temporary and perishing; and where they proceed not from some *cause* in the character and disposition of the person who performed them, they can neither redound to his honour, if good; nor infamy, if evil. The actions themselves may be blameable; they may be contrary to all the rules of morality and religion: But the person is not answerable for them; and as they proceeded from nothing in him, that is durable and constant, and leave nothing of that nature behind them, it is impossible that he can, on their account, become the object of punishment or vengeance [section VIII, part ii].

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

moderate comparative claim contained in the Fundamental Attribution Error, then it may not support skepticism about character and so may not threaten characterological conceptions of responsibility. Even if situationism does support skepticism about global character traits, we might try to reformulate characterological conceptions of responsibility so that they do not presuppose cross-situational consistency.

However, there are other serious worries about characterological conceptions of responsibility that are completely independent of situationism. The more serious worry is that characterological conceptions are both under-inclusive and over-inclusive about who is responsible and blameworthy. A characterological conception promises to be under-inclusive about whom to blame or punish, because it will exculpate people who are reasons-responsive and choose to do wrong, provided that choice is out of character. For instance, it will exculpate Richard Herrin for his deliberate murder of Bonnie Garland, on the ground that this was out of character, despite the fact that there was no evidence that he suffered from significant cognitive or volitional impairment.[30] The characterological conception of responsibility promises to be over-inclusive about whom to blame or punish as well, because it will not exculpate for wrongdoing that is in character, even if the agent's character lacks elementary capacities for reason-responsiveness.

Doris himself wants to integrate situationism with an identificationist conception of responsibility. Harry Frankfurt introduced the identificationist idea, which others have endorsed with various modifications.[31] The identificationist proposal says, as a rough first approximation, that one is responsible for one's action when one endorses or identifies with one's determining motive. Most agents identify with their determining motives, wanting to want what they want or regarding what they want as valuable. But in some cases agents may not approve or endorse what they want, as in the case of the unwilling drug addict. Perhaps many of the subjects in the Milgram or Stanford Prison experiments did not endorse their desires and actions. I do not think the experimental studies are clear about this, but it is possible. Would this mean that the subjects in those studies were not responsible for their behavior and should be excused?

We should be skeptical of identificationist conceptions of responsibility. For one thing, they have a hard time holding agents who are weak-willed responsible for their actions. For in the case of weakness of will or akrasia, the agent acts contrary to what she judges best, because she is moved by good-independent appetite or desire. But then the akratic actor typically does not endorse or approve the desires that in fact determine her will

[30] The Herrin case is discussed, as part of a similar worry about the under-inclusiveness of characterological conceptions of responsibility, in Moore, *Placing Blame,* 578–80.

[31] See Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5–20 and Gary Watson, "Free Agency," *Journal of Philosophy* 72 (1975): 205–20.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

when she is weak-willed. But we normally hold akratic actors responsible and do not in general treat weakness of will as an excuse.

In fact, given the worries about the under-inclusiveness and over-inclusiveness of characterological conceptions, which Doris seems to recognize (pp. 130–32), it is a little surprising that he wants to embrace an identificationist conception of responsibility. For it, like the characterological conception, threatens to be both under-inclusive and over-inclusive. It will be underinclusive, because it will excuse when someone does not identify with her guiding motive, even if she was normatively competent for her choice of wrongdoing. Perhaps Richard Herrin did not approve of his homicidal motives, but if he was competent and intended to murder Bonnie Garland, then he is responsible. It will be over-inclusive, because it will count as culpable, wrongdoing that the agent endorses, even if she was not reasons-responsive at the time. Doris's introduction of a constraint of *narrative integration* into the identificationist proposal is supposed to make it less under-inclusive, because it will not excuse for failure to approve of one's determining motive if those motives are nonetheless part of a larger plan or narrative that guides one's life (pp. 142–46). But it is far from clear that narrative integration solves these problems. Indeed, like the characterological view, the narrative view invokes patterns in one's choices and behavior. But with the characterological view it was precisely the idea that anomaly, as such, was excusing that led to the under-inclusiveness worry. So it is hard to see how narrative integration will not make the resulting identificationist view under-inclusive as well.

I have articulated these worries about both characterological and identificationist views against the backdrop of a conception of responsibility as fair opportunity to avoid wrongdoing that requires intentional choice by an agent who is normatively competent and has situational control. Doris does not recognize this exact conception of responsibility, but he does discuss normative competence conceptions of responsibility and suggests that situationism poses serious challenges for such conceptions.

Before turning to the bearing of situationism on normative competence, it might be worth pausing over the implications of situationism for situational control. For one might have thought that this would be the aspect of responsibility that situationism challenged. Perhaps we only think that we possess situational control while we are in the grips of the Fundamental Attribution Error. Once we recognize the pervasive influence of situational factors on our behavior, perhaps we should conclude that situational control is an illusion or at least frequently impaired.

However, it is hard to see factors of the sort studied by situationists as compromising situational control. As we have seen, the paradigmatic excuse that denies situational control is duress. But in the criminal law duress requires wrongful coercion or threats by another agent that leaves no reasonable alternative to wrongdoing that complies with

the threat. This standard is clearly not met in any of the situationist studies we have discussed. In many cases, there is no other agent intervening to apply pressure on the wrongdoers. The closest we come to coercion is the Milgram experiment in which the experimenter requests the subject to shock the learner to correct his mistakes. But there is no reasonable interpretation of coercion in which such a request is coercive, and it certainly leaves open reasonable and permissible alternatives to compliance with the request. Even if we drop the criminal law's demand that duress have its source in the coercive intervention of another agent and allow duress excuses in response to hard choices, whatever the cause, the situationist cases still will not involve duress. For the choice agents face when they engage in wrongful behavior in these studies is just not sufficiently hard choice, and they all have reasonable and permissible alternatives to wrongdoing.

Because there is no plausible duress excuse for bad behavior in the situationist studies, we might instead see if situational factors compromise the agent's normative competence. This is the claim that Doris develops (pp. 133–34). He begins by claiming that situational factors compromise self-control and that they show that self-control is more the exception than the rule (p. 134). But self-control is involved in normative competence, and so Doris concludes that situationism challenges the cognitive and volitional dimensions of normative competence (p. 136). He claims that situational factors may impair one's ability to recognize reasons for action—perhaps the hurried seminarians do not even notice the slumping confederate—or may make it difficult for one to sustain and implement one's normative judgments—perhaps group effects compromise volitional capacities, and Milgram subjects are disabled from acting on their moral knowledge by their desire to comply with authority (pp. 136–38).

But most 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 subconscious, to conform to the firm requests of the experimenter. But while many people did in fact act surprisingly badly, there is nothing in the situationist gloss on 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 of *competence.*

It seems extremely plausible that these situational factors affect performance, not competence. After all, the situationist paradigm precisely targets capacities and traits that we take to be commonplace, rather than exceptional or heroic. A significant number of subjects manifest this capacity in expected behavior, and, even those who do not, show signs of discomfort and stress that evidence the existence of the capacity for appropriate action.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

Life is replete with unexercised capacities of various kinds. One example is John Austin's case of the putt he could have made but didn't.[32] Another example is the dieter's weakness for a dessert that he should and could have avoided. In analyzing these ascriptions of capacities, we are positing an actual psychological structure, but one which funds relevant counterfactual claims about how the agent would have behaved in other circumstances. If determinism is true, we do not think that the agent would behave any differently in circumstances *exactly* like the actual circumstances. Instead, we think that he could have acted differently in the actual circumstance if he would have acted differently in circumstances that were relevantly different.[33] Of course, relevant differences cannot be too great; otherwise, we will not be tracking actual capacities. We cannot test whether I could have made the putt that I failed to make by asking whether I would have made the putt if I first became as good a golfer as Tiger Woods. Exactly how to constrain the relevant counterfactuals is a complex matter, but it is reasonably clear that the relevant differences should be familiar and modest. And one counterfactual does not confirm or disconfirm a capacity. Incompetent golfers can make lucky putts once in a great while, and Tiger Woods occasionally misses putts that are makeable for him. What we should expect from someone who has a capacity is that in the relevant counterfactuals he will perform successfully on a *regular* basis. But by this test it seems plausible that most subjects that did not display compassionate behavior could have.

Consider the Milgram experiment again. Would the compliant subjects have administered serious shocks if they had been given more time to consider their options, if they had been asked to justify the imposition of apparent harm to innocent parties, or if they had known the learners? If subjects would not have shocked in some of these circumstances, that provides evidence that those subjects had a capacity for recognizing reasons for compassion and for engaging in compassionate behavior.

Interestingly, several variations that Milgram himself devised in his basic experimental design reinforce these claims. Though the variations involve different subjects, they provide evidence about subjects in general. That is, the counterfactuals that are most relevant to whether a given agent had a capacity to behave differently than she in fact did act in actual circumstances ask how *she* would have behaved in relevantly different circumstances. But if we cannot test how she would have behaved in such circumstances, we can nonetheless appeal to how other people do behave in such circumstances. Unless we have reason to believe that her capacities are unusual in this regard, the

[32] John Austin, "Ifs and Cans," in John Austin, *Philosophical Papers* (Oxford: Clarendon Press, 1961), 166.

[33] See Daniel Dennett, *Freedom Evolves* (New York: Viking, 2003), 75–76.

behavior of others in relevant circumstances can be defeasible evidence about how she would behave and what capacities she has. In a variation in which subjects had to place the learner's hand on an electronic plate in order for him to receive the (apparent) shock only 30 percent of subjects administered the maximal shock, as compared with 62.5 percent in the original study.[34] In a later variation in which the experimenter left the room and gave instructions remotely, only 20 percent of subjects administered the maximal shock, even though this required the subject to lie to the experimenter about the shocks she administered.[35] Importantly, in a variation in which the subjects themselves determined the severity of the shock, 95 percent of subjects refused to administer shocks beyond the point at which the learner protested.[36]

These are reasons to think that appropriate counterfactual evidence would vindicate ordinary assumptions that most people have the capacity for recognizing and responding to reasons for compassionate behavior, even when they do not exercise these capacities. If situationist studies reveal facts about performance, rather than competence, then situationism does not compromise the sort of normative competence necessary for responsibility.

Of course, it may be that in some contexts situational factors genuinely do compromise cognitive or volitional aspects of normative competence. For instance, as we will see (in Sections VI and VII below), there is a good case to be made that the normative competence of soldiers in the heat of battle is potentially compromised in several ways—by making high stakes decisions in split seconds, often under significant uncertainty in circumstances in which their passions and instincts can be inflamed. The point is not that situationist factors cannot affect normative competence but rather that situational factors do not, as such, compromise normative competence.

Doris suggests that situationism poses another "more insidious" problem for normative competence (p. 138). Because people make the Fundamental Attribution Error, they are often ignorant of the influence of situational factors on their decision-making. To this extent, they do not understand their own cognitive processes (p. 139). But fair opportunity conceptions of responsibility do not insist that psychological transparency is necessary for normative competence. They claim only that agents are responsible just in case they engage in wrongful conduct and were normatively competent at the time. Agents may often act wrongfully under the influence of situational factors that they do not fully understand, but that does not show that they lack the capacity to do otherwise.

[34] Milgram, *Obedience to Authority*, 34–35.
[35] Ibid., 59–60.
[36] Ibid., 70–72.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

VI. Situationism and Wartime Wrongdoing

Even if situationism does not in general compromise responsibility, it might be that the situationist perspective on particular situations suggests a significant impairment of normative competence or situational control. In "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity" Doris and Murphy make this sort of more restricted argument when they argue that ideas from situationist psychology support excusing a large variety of wartime wrongdoing, including not only crimes committed by combatants, such as the killing of noncombatants and the use of disproportionate force, but also the commission of noncombat-related crimes committed during wartime by soldiers, such as crimes committed by guards in military prisons.[37] They do not frame their argument in terms of a particular conception of responsibility and excuse, but they do make appeals to the importance of normative competence, which is the relevant aspect of our conception of responsibility as the fair opportunity to avoid wrongdoing.

Their argument appeals to both *distal* and *proximate* factors affecting the normative competence of soldiers under stress. They mention three distal influences (pp. 38–41).

- Military training reduces inhibitions on using violence.
- Military training reduces independence of judgment and increases respect for authority and following the chain of command.
- In various ways, military training demonizes the enemy, often representing them as morally inferior or sub-human.

But we must remember that military training also teaches soldiers to respect various constraints in war such as the requirement to target only combatants, to use proportionate force in self-defense, and to minimize collateral harm. This is one reason it is important to recognize, in addition to these distal influences on soldiers, several proximate factors potentially affecting the normative competence of soldiers in battle. Perhaps the most obvious proximate factor, which they do not mention, is that often battlefield decisions have to be made instantaneously without the benefit of careful and dispassionate assessment. They cite other proximate factors that tend to corrupt judgment (pp. 35–38).

- The diffusion of responsibility that individuals experience in groups.
- The onslaught of noise and other perceptual stimuli.
- The filth and odors in battle that confound the senses and lead to a sense of pollution and contamination.

[37] I provide references to Doris and Murphy, "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity" parenthetically in the text hereafter.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

- Exhaustion and lack of sleep.
- The emotional toll of losing comrades to injury and death.

Doris and Murphy claim that the combination of distal and proximate factors produces "moral drift" and impairs normative competence.[38]

When all or most of the distal and proximate factors are present, there is a plausible basis for a full or partial excuse for some forms of wrongdoing. Their argument is strongest for wrongdoing committed by combatants in the heat of battle. Call this *hot combat.* For instance, military training prior to combat treats independence of judgment as a vice and obedience to the chain of command as a virtue. Moreover, the immediate context of battle calls for quick decision-making and contains a barrage of stimulus, both of which may compromise decision-making capacities that are central to normative competence (pp. 35–41). It will often be hard under battle conditions to know what is right and wrong, because it is hard to distinguish between combatants and noncombatants or to assess proportionate responses instantaneously and under chaotic conditions. And even if one can recognize what duty requires, it may be very hard to resist orders to act differently or to resist the pressures of comrades to participate in wrongdoing.

It is not clear whether wrongdoing under such circumstances should be afforded a full or partial excuse, that is, a genuine defense or mitigation. There is no reason to expect a general answer. Verdicts would have to be case-specific. One thing that is especially interesting about this kind of excuse for wartime wrongdoing is that it illustrates ways in which situational factors can compromise normative competence. This is true whether the excuse provided is full or partial.

Notice two important limitations in this rationale for excuse. First, it depends heavily on the proximate factors present in hot combat. Second, even in cases of wrongdoing committed in the heat of battle and pursuant to the orders of superiors, there are limitations on what can be excused. In particular, even compliance with the orders of a superior does not excuse *manifest illegality.*[39] Manifest illegalities are actions that are wrong on their face and clearly exceed the bounds of justified force. Manifest illegalities would include the intentional targeting of noncombatants, the torture and execution of prisoners, and the rape of noncombatants or combatants.

However, Doris and Murphy are surprisingly promiscuous with excuses, extending excuses from these core cases in which understandable forms

[38] The military criminal code appears to rely on a M'Naghten-style cognitive conception of normative competence. See *Manual for Courts Martial* 2012, Rule 706, especially 706(c)(2)(A), and 916(b)(2). My own view is that the failure to recognize a distinct volitional dimension to normative competence is a mistake here, as it is in the M'Naghten test.

[39] See *Model Penal Code* section 2.10 and *Manual for Courts Martial* 2012, Rule 916(d).

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

of wrongdoing are committed in the heat of battle to various kinds of peripheral cases.

> In particular, we need to say something about how far from the battlefield the kinds of excusing conditions we identify reach, and what behaviors they extend to. It might be thought that the scope of our argument is actually quite limited, covering only certain sorts of "unnecessary killing" during combat operations, and that the further from these central cases we look, the less plausible the argument becomes. This is certainly a reasonable position to adopt, and we agree that the further one strays from the manifestly *legal* killing of enemy soldiers actively engaged in combat, the more difficult it is to establish the presence of excusing conditions. However, we think that the scope of our argument is surprisingly broad . . . [p. 46].

Indeed, they think that the argument for excuse extends from hot combat to cases involving wrongdoing committed by soldiers outside of immediate combat situations. We might call these cases of *warm combat* in which soldiers who do spend time in combat commit wrongdoing in wartime but outside of immediate combat. One such case is the killing of at least four hundred unarmed noncombatants by members of Charlie Company during the Vietnam War in 1968.[40] Most of the victims were women and children, some of whom were raped before being killed. Another such case is the killing of 24 unarmed civilians by U.S. Marines in Haditha in the Al Anbar province of Iraq in 2005. In both cases, the killings were widely held to be committed in retaliation for recently suffered combat casualties.

Doris and Murphy want to extend further the excuse for wartime wrongdoing to cases of *cold combat* involving soldiers away from battle acting as prison guards. The most visible recent such case involved the physical, psychological, and sexual abuse and torture of prisoners by military personnel in the U.S. Army at the Abu Ghraib Prison in Iraq in 2004.[41] Surprisingly, Doris and Murphy do not invoke the Stanford Prison Experiment in their argument for excuse at Abu Ghraib.[42] We might think that just as otherwise normal young men can become abusive toward others when given authority over them, so too can young men and women when they are made real, and not just pretend, prison guards with real, and not just pretend, authority. Indeed, we might expect that such abuse is all the more likely when practiced by members of the military, who

[40] See Seymour Hersh, "The My Lai Massacre," a three-part article in the *St. Louis Dispatch,* November 1969 and Michael Bilton and Kevin Sim, *Four Hours in My Lai* (New York: Viking, 1992).

[41] Seymour Hersh, "Torture at Abu Ghraib," *The New Yorker* May 10, 2004.

[42] By contrast, in *The Lucifer Effect,* Zombardo explicitly notes and explores parallels between the Stanford Prison Experiment and wartime atrocities, especially at Abu Ghraib.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

have become desensitized to suffering, against real prisoners thought to be enemy combatants. Reflecting on the Stanford Prison Experiments, we might be tempted to think: There, but for the grace of God, go I.[43]

Moreover, though they recognize the distinction between full and partial excuses or mitigation, they nonetheless argue that wartime atrocity should typically receive a full excuse, and not just mitigation (p. 49). Furthermore, they think that in wartime it becomes extremely difficult to distinguish between manifest illegality and military necessity (pp. 42, 45). Doris and Murphy do not claim that all cases of wartime wrongdoing should be excused, but they do claim that the presumption should be in favor of excuse, rather than culpability (pp. 26, 46).

VII. Excusing War Crimes?

War crimes raise an interesting and important issue in applied responsibility, and it seems quite plausible that normative competence is sometimes compromised in hot combat, providing an excuse, whether full or partial, for wrongdoing committed in such circumstances. We should be sympathetic toward soldiers who have to make life and death decisions in an instant amid a barrage of confusing informational and sensory stimuli. Making fine moral discriminations in such circumstances can be like trying to drink from a fire hose. We should be prepared to excuse difficulty soldiers might have distinguishing combatants from noncombatants and from assessing what is proportionate force to be used in disabling potential aggressors. However, the full scope of Doris and Murphy's excuse for wartime wrongdoing is quite extraordinary in several respects.

For one thing, it is quite surprising that Doris and Murphy should suppose that there should be a presumption in favor of excuse for wartime wrongdoing of various kinds. In the criminal law, the prosecution has the burden of proof for establishing the elements of the crime—*actus reus* and *mens rea*—beyond a reasonable doubt. But with affirmative defenses, which concede the elements of the offense but allege justification or excuse, it is the defense that has the burden of proof by a preponderance of the evidence or by clear and convincing evidence. So, in particular, it is the burden of the wrongdoer to show that he is entitled to an excuse, because either his normative competence or his situational control was significantly impaired. In part because defenses of justification and excuse are designed to be exceptional, it makes sense that those asserting such defenses should bear the burden of proof.

Moreover, it is peculiar that Doris and Murphy acknowledge that the rationale for the excuse in hot combat depends on distal and proximate

[43] Those not religiously inclined may think instead: There, but for the natural and social lotteries, go I.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

factors that are either absent or greatly reduced in situations of warm combat and virtually nonexistent in situations of cold combat. They recognize a distinction between full excuse and mitigation (or partial excuse) and recognize that excusing factors are scalar but nonetheless insist on defending full excuse in a wide range of cases. Since the plausibility of a full excuse in the case of hot combat depends crucially on the heat of battle, why not think that the case for excuse outside the heat of battle is significantly reduced? Furthermore, the killings in My Lai and Haditha and the torture and abuse at Abu Ghraib seem to include clear cases of manifest illegality that would fall outside the scope of excuse even in hot combat. The basis for excuse cannot be reduced or lost without losing or reducing the excuse.

Consistent with Doris's argument in *Lack of Character,* Doris and Murphy conflate issues of competence and performance. Responsibility requires normative competence, and normative competence requires capacities to make normative discriminations and act on those normative judgments. Since responsibility requires these capacities, it is not compromised whenever they are not properly exercised. Indeed, if they were, we could never hold people responsible for wrongdoing. But situationism only says something about patterns in people's actual behavior, how situational factors tend to influence behavior. These facts about performance in no way indicate an inability to reason or intend otherwise. The same is true in war. Situational evidence, by itself, only shows things about how soldiers behave and the influence of situational factors on their conduct. It does not demonstrate incapacity. In some cases involving hot combat, this further claim might be warranted. It might well be that in situations involving risk of serious personal harm and death in which decisions have to be made in an instant and in which there is a barrage of sensory stimuli, the ability to tell right from wrong and to implement the right decision despite distraction and countervailing passion is severely compromised. But there is no reason to assume that this competence is seriously compromised in warm and cold combat situations. All Doris and Murphy offer by way of evidence for the extension is situationist claims about how common moral drift is even in such situations. But unless we are to excuse all wrongdoing, we have to distinguish between performance and competence and demand evidence of genuine incapacity.

Their support for a widespread excuse for wrongdoing in wartime should strike us as promiscuous about excuse and insufficiently serious about accountability. They try to meet these worries by adopting the surprising position that a full excuse does not preclude criminal liability. In effect, they urge that military tribunals recognize strict liability. That is, they conclude that we should excuse the wrongdoers but punish them anyway! This is an extraordinary claim. For one thing, a broadly retributivist conception of blame and punishment, according to which blame and punishment are deserved in proportion to culpable wrongdoing, is

very plausible and a deep feature of the criminal law. Denying either wrongdoing or culpability rebuts the case for blame and punishment. Whereas justifications deny wrongdoing, excuses deny culpability or responsibility. So a full excuse that denies culpability or responsibility altogether precludes guilt and punishment. This retributivist argument is reflected in the criminal law insofar as excuses are affirmative defenses, which, if successful, justify acquittal. Moreover, ensuring the fair opportunity to avoid wrongdoing renders strict liability in the criminal law deeply problematic. It is possible but controversial to have strict liability crimes that do not require intent, knowledge, recklessness, or negligence as part of *mens rea* (an ingredient in the offense itself), but even crimes that are strict liability in this sense admit of an excuse in terms of insanity or diminished competence. There are no strict liability crimes that do not admit of excuse because of impairments to normative competence or situational control. As excuses, these conditions block blame and punishment. If we were to accept Doris and Murphy's argument that most wartime wrongdoing should be excused, we would have to forego blame and punishment. The difficulty of accepting that conclusion is not reason to embrace strict liability for war crimes but rather reason to have a more discriminating conception of excuse.

Part of having a more discriminating conception of excuse involves taking normative competence seriously and not excusing wartime wrongdoing in which the soldier's cognitive and volitional capacities are not significantly compromised. Another part of a more discriminating conception of excuse might be to recognize the possibility of more fine-grained responses to wartime wrongdoing. Normative competence and, hence, responsibility are scalar phenomena, even if we sometimes employ thresholds and bright lines in our decisions about whom to hold responsible and whom to excuse. While a full excuse for wartime wrongdoing should presumably be exceptional and limited to wrongs that are not manifestly illegal and are committed in the heat of battle or under direct orders from a superior, not all combat situations compromise a soldier's normative capacities to the same degree. Consequently, we may have reason to recognize *partial* excuses even when a full excuse cannot be justified. This is fairly easily accommodated within our beliefs about moral responsibility and blame, because we already recognize that moral responsibility can be scalar. However, American criminal law does not in general recognize a doctrine of partial excuse.[44] But even in systems that do not recognize partial excuse, the same facts about diminished capacity that would otherwise justify a partial excuse can and should operate as

[44] The one exception to the generalization that American criminal law does not recognize partial excuse is the provocation defense, under which intentional homicides committed with adequate provocation reduce to an offense of voluntary manslaughter. Even this is not a clear exception inasmuch as it is unclear whether provocation is best conceptualized as a partial justification or a partial excuse.

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press

mitigating factors at sentencing. A further resource for fine-tuning blame and punishment in situations of partial responsibility is the distinction between *censure* and *sanction.* The criminal law typically both censures and sanctions culpable wrongdoing. But these responses can come apart. In situations in which it may have been hard for a soldier to exercise good judgment that did not sufficiently impair her normative competence to qualify for a full excuse, we may nonetheless consider mitigation or the possibility of censuring her and her conduct without imposing sanctions.

Philosophy, University of California, San Diego

https://doi.org/10.1017/S026505251300006X Published online by Cambridge University Press