

To appear in *Oxford Studies in Agency and Responsibility* 1 (2013).

FAIRNESS AND THE ARCHITECTURE OF RESPONSIBILITY¹

David O. Brink and Dana K. Nelkin

In this essay, we explore a conception of the nature and structure of responsibility that draws on ideas about moral and criminal responsibility. Though the two sorts of responsibility are not the same, the criminal law reflects central assumptions about moral responsibility, and the two concepts of responsibility have very similar structure. Our conception of responsibility draws on work of philosophers in the compatibilist tradition who focus on the choices of agents who are reasons-responsive and work in criminal jurisprudence that understands responsibility in terms of the choices of agents who have capacities for practical reason and whose situation affords them the fair opportunity to avoid wrongdoing.² We treat these two perspectives as potentially complementary and argue that each can learn things from the other. Specifically, we think that criminal jurisprudence needs a more systematic conception of the capacities for normative competence and that ideas from the reasons-responsive literature on moral responsibility, some familiar and some novel, can fill this need. However, we think that moral philosophers tend to focus on the capacities involved in responsibility and so tend to ignore the situational element in responsibility recognized in the criminal law literature. Our conception of responsibility brings together the dimensions of normative competence and situational control, and we factor normative competence into cognitive and volitional capacities, which we treat as equally important to normative competence and, ultimately, responsibility. Moreover, we argue that normative competence and situational control can and should be understood as expressing a common concern that blame and punishment presuppose that the agent had a fair opportunity to avoid wrongdoing. Thus, we treat the value that criminal law theorists associate with the situational element of responsibility as the umbrella concept for our conception of responsibility, one that explains the distinctive architecture of responsibility.

¹This essay is fully collaborative. The authors are listed in alphabetical order. The ideas for this essay grew out of a graduate seminar that we taught together on the topic of partial responsibility in 2008 and were refined in a seminar on responsibility that DB taught in 2011. Versions of this material were presented at the University of Illinois, the University of Calgary, the Murphy Institute at Tulane University, Cornell University, the University of Western Ontario, and the New Orleans Workshop on Agency and Responsibility. We would like to thank audiences on those occasions for helpful feedback. We owe special thanks to input from Craig Agule, Sarah Aikin, Amy Berg, Mitch Berman, Michael McKenna, Per Milam, Richard Miller, Michael Moore, Stephen Morse, Derk Pereboom, Erick Ramirez, Sam Rickless, Tim Scanlon, David Shoemaker, Jada Twedt Strabbing, Sarah Stroud, Matt Talbert, Michael Tiboris, and Gary Watson.

²The philosophical work in the reasons-responsive wing of the compatibilist tradition on which we draw includes Fischer and Ravizza 1998, Wallace 1994, Wolf 1990, and Nelkin 2011. The criminal jurisprudence work on which we draw includes Hart 1957 and 1961; Moore 1997; and Morse 1994, 2002, and 2003.

This essay aims to motivate and articulate this sort of fair opportunity conception of the architecture of responsibility. It is part of a larger project that develops this conception and applies it to issues of partial responsibility, involving insanity and psychopathy, immaturity, addiction, provocation, and duress. The details and applications of the fair opportunity conception of responsibility are interesting and important, and we hope to address them more fully elsewhere. But the framework itself is important and requires articulation.

1. RESPONSIBILITY, BLAME, AND THE REACTIVE ATTITUDES

P.F. Strawson famously highlighted the link between ascriptions of responsibility and the reactive attitudes (Strawson 1962). The reactive attitudes involve emotional responses directed at oneself or another in response to that person's conduct. Reactive attitudes include hate, love, pride, gratitude, anger, regret, resentment, indignation, and forgiveness. So understood, the reactive attitudes form a large and heterogeneous class. Some of these reactive attitudes have little direct connection with moral praise and blame and responsibility. Consider the difference between anger and resentment. Anger need not have moral content. I might be momentarily angry or upset with a very young child who has carelessly damaged a treasured keepsake of mine. But resentment would seem to be out of order. Resentment seems to involve a kind of anger or upset that presupposes that one has been mistreated or wronged by another in some way. This kind of moral judgment does not seem to apply to a very young child. The idea that assumptions about responsibility are embedded in the reactive attitudes makes most sense if we focus on this narrower class of reactive attitudes that are *moralized* (cf. Wallace 2004: ch. 2).

In particular, we want to focus on the attitudes and practices of praise and blame, especially as they reflect assumptions about responsibility. Some negative reactive attitudes, such as regret, don't seem to implicate responsibility at all. Bernard Williams describes the case of a truck driver who, through no fault of his own, hits and kills a child who has darted into the street (1976: 28). As Williams claims, it is appropriate for the driver to feel a kind of *agent-regret* at being the instrument of the child's death, which is distinct both from the regret or horror that bystanders might feel and from guilt for having been responsible for wrongdoing.

Normally, blame only makes sense if the agent is responsible for some kind of wrong. Some philosophers distinguish between two kinds of blame and responsibility. For instance, Gary Watson distinguishes between responsibility as *attributability* and as *accountability* (1996). An agent is responsible in the attributive sense, roughly speaking, when her actions reflect *the quality of her will* in the right way. Some kinds of blame can be a fitting response to the quality of the agent's will. For instance, A might have hard feelings toward B if B injures A through malice, recklessness, or negligence. Here, our reactive attitudes track the insufficient regard that B shows A's interests and rights. But attributability does not guarantee accountability. Consider a situation in which we find out that though B's actions exhibit malice, in no relevant sense did B have an opportunity to do otherwise, perhaps because he suffers from a serious mental illness and is not a competent decision-maker as a result. In these cases, we are likely to think that B was not at *fault* or *culpable* and so not *accountable* for the harm he did. Although hard feelings may remain and be perfectly appropriate in such cases, reactive attitudes involving resentment and indignation cease to be appropriate and tend to dissipate. Attributability is necessary, but

not sufficient for accountability. In this essay, we are especially interested in responsibility as accountability and its connection with reactive practices and attitudes involving blame, and we rely on an intuitive understanding of the reactive attitudes that seem to be especially responsive to accountability.³ It is this sense of blame and responsibility that we take to be most relevant to the sort of responsibility required for punishment and to capture what is common to both moral and legal responsibility.

Thus, our focus in what follows will be restricted to attitudes of praise and blame involving accountability, with special attention to attitudes and practices of blame, rather than praise. We do so because our aim is to combine insights from criminal jurisprudence, which focuses on criminal acts that involve wrongdoing, with those from moral theory. But we believe that there are natural ways of extending what we say here about attitudes of blame and blameworthy actions to praise and praiseworthy actions.

Strawson links responsibility and reactive attitudes, such as those of resentment and indignation, in a biconditional fashion.

Reactive attitudes involving blame and praise are appropriate just in case the targets of these attitudes are responsible.

Call this biconditional claim *Strawson's thesis*. Strawson's thesis can be interpreted in two very different ways, depending on which half of the biconditional has explanatory priority.

According to the first interpretation, there is no external, or response-independent justification of our attributions of responsibility. This reading fits with Strawson's view that our reactive attitudes and ascriptions of responsibility, as a whole, do not admit of external justification. Particular expressions of a reactive attitude might be corrigible as inconsistent with a pattern of response, but the patterns of response are not themselves corrigible in light of any other standard. Similarly, particular ascriptions of responsibility might be corrigible in light of patterns in our ascriptions of responsibility, but the patterns themselves are not corrigible in light of any other standards. Responsibility judgments simply reflect those dispositions to respond to others that are constitutive of various kinds of interpersonal relationships. This is a *response-dependent* interpretation of Strawson's thesis.

This response-dependent interpretation of Strawson's thesis is probably the right interpretation of Strawson.⁴ But as a systematic, rather than an interpretive, matter, we favor an alternative interpretation of Strawson's thesis that is *realist*, rather than response-dependent. This interpretation stresses the way that the reactive attitudes make sense in light of and so *presuppose* responsibility. As such, the reactive attitudes are *evidence* about when to hold people responsible, but not something that constitutes them being responsible. It's true that the reactive attitudes are appropriate if and only if the targets

³Here, we are in agreement with Watson (1996: 276, 2011). By contrast, T.M. Scanlon develops a conception of blame that seems to presuppose only attributability, not accountability (2008: ch. 4, esp. p. 202).

⁴Watson defends this response-dependent interpretation of Strawson's thesis, at least on interpretive grounds (1987: esp. 222). Wallace defends this interpretation of Strawson's thesis in its own right (2004: 19), though we think other elements in Wallace's account fit better with an alternative realist reading.

are responsible, but it's the responsibility of the targets that makes the reactive attitudes toward them fitting or appropriate. In the biconditional relationship between responsibility and the reactive attitudes, it is responsibility that is explanatorily prior, according to this realist interpretation. Strawson points out that the limits of our reactive attitudes are indicated by our practices of exemption and excuse. Because the realist believes that the reactive attitudes presuppose responsibility, she can appeal to our practices of exemption and excuse to help understand the conditions under which we are responsible. This will be a response-independent conception of responsibility.

A response-independent conception of responsibility is hostage to traditional worries about freedom of the will. The problem of free will is the problem of reconciling responsibility with determinism, because responsibility may seem to presuppose freedom of the will, and freedom of the will may seem incompatible with determinism. Our realist approach to responsibility and the reactive attitudes is best articulated as a version of compatibilism that denies that responsibility requires a form of freedom that would be undermined by the truth of determinism. In particular, because our practices of exemption and excuse track forms of normative competence and situational control, rather than the truth of determinism, they promise to ground a compatibilist conception of responsibility. Though we will articulate this compatibilist interpretation of our project, we cannot defend it here (though we return to these issues briefly in §7 below).⁵

Although Strawson focuses primarily on the relation between the reactive attitudes and responsibility, his thesis fits well with a particular approach to punishment and criminal responsibility. In particular, the realist interpretation of Strawson's thesis fits with a broadly retributive approach to blame and punishment, precisely because the retributivist thinks that the reactive attitudes and our practices of blame and punishment can be appropriate responses to culpable wrongdoing, where culpable wrongdoing is wrongdoing for which the agent is responsible. To see this, it will be useful for us to say more about both blame and punishment.

When agents are responsible (accountable) for doing wrong blame is appropriate. Blame typically involves both *censure* and *sanction*. When we blame someone, we not only censure her conduct but also censure the agent herself for engaging in that conduct. Parents are often warned to disapprove the bad conduct of their children but not to blame them. This is because blame involves finding fault in the agent and that seems to assume that the agent is responsible (accountable) and could have avoided the conduct. Being blameworthy licenses various kinds of sanction, often informal and sometimes formal. Blame itself can involve overt reproach, which is a kind of sanction, whether directed at

⁵Because we think that reactive attitudes involving praise and blame presuppose that the targets of these attitudes are responsible, we accept the need to provide a response-independent conception of responsibility and to answer skeptical doubts about responsibility. Consequently, we see response-dependent conceptions of responsibility as offering skeptical solutions to skeptical worries (cf. Kripke 1982: 66-67). We view skeptical solutions to skeptical problems as, at best, a kind of fallback solution to be entertained only after straight solutions have clearly failed. For a fuller exploration of the compatibilist aspects of this conception of responsibility, see Nelkin 2011. We believe that at least some of what we say here can be accepted by incompatibilists who accept *further* conditions on responsibility, beyond that of indeterminism.

another or at oneself. Sometimes reproach is the only appropriate sanction. But sometimes blameworthiness licenses other informal sanctions, such as public rebuke or social distancing. And in other cases, blameworthiness might license various kinds of punishment, whether personal, social, or legal. To be blameworthy is to be a fitting object of blame, censure, and sanction. It is to be deserving of these attitudes and responses. No doubt, where sanctions are appropriate, they have to be proportionate, and there may be cases in which one is blameworthy and yet it is not on balance appropriate to blame or sanction. But, presumably, even in these cases there is a *pro tanto* case for blame and at least some informal sanction, if only self-reproach, as a fitting response to culpable wrongdoing.

On this view, punishment is a species of blame for culpable wrongdoing. On a broadly retributive view of the criminal law, this is true of legal punishment as well. We understand criminal punishment as the authorized deprivation of an agent's normal rights and privileges, because he or she has been found guilty of a criminal act.⁶ Punishment is a form of blame, and like other kinds of blame, presupposes culpable wrongdoing. Legal retributivism, as we understand it, is the claim that legal punishment is justified on the basis of culpable legal wrongdoing. This claim can take *positive* or *negative* forms. According to positive retributivism, culpable wrongdoing is both necessary and sufficient for justifying punishment. The sufficiency claim admits of both *strong* and *weak* interpretations. According to strong sufficiency, culpable wrongdoing is a sufficient condition of justified proportional blame and punishment, whereas, according to weak sufficiency, culpable wrongdoing is sufficient for a *pro tanto* case for proportional blame and punishment. Weak sufficiency allows for the *pro tanto* case for retributive blame and punishment to be overridden in particular cases by non-culpability moral considerations, such as forgiveness or mercy. By contrast, according to negative retributivism, culpable wrongdoing is necessary, but not sufficient, for justified punishment.⁷

Legal retributivism (in either version) has the virtue of explaining well the two principal forms of affirmative defense in the criminal law. According to the retributivist, justified punishment aspires to track culpable wrongdoing.⁸ Wrongdoing and culpability are independent variables. Affirmative defenses, whose success justifies acquittal, deny either wrongdoing or culpability. *Justifications*, such as the necessity defense, deny wrongdoing, insisting that behavior that would otherwise be wrong is not in fact wrong in these circumstances. *Excuses*, such as the insanity defense, deny culpability or

⁶Cf. Bedau and Kelly 2010. We aim for a normatively neutral and ecumenical definition of punishment and one that identifies punishment as involving deprivations of certain sorts, but not essentially involving the imposition of pain or suffering.

⁷Cf. Duff 2008. The view that we have called "negative retributivism" is sometimes called a "mixed theory" of punishment, because it requires more than one type of justificatory reason, typically, both retributivist and consequentialist. It is also worth noting that our definition of retributivism does not commit retributivists to endorsing the thesis that punishment is intrinsically good, as some retributivists claim.

⁸Precisely for this reason, a skeptic about moral responsibility will deny that any retributivist view of punishment can be correct. For a skeptical view and its relation to punishment, see Pereboom 2012.

responsibility, claiming that the agent acted wrongly but was not responsible for her wrongdoing.

Here, the criminal law reflects the moral landscape well. Moral retributivism could be understood as the claim that moral blame (that presupposes accountability) and informal sanction are appropriate only as a response to culpable moral wrongdoing. It too has the virtue of explaining the two principal ways of avoiding blame -- justifying and excusing conduct. Justification denies wrongdoing, and excuse denies responsibility for wrongdoing. Insofar as moral retributivism says that moral blame ought to track desert, where desert is the product of the two independent variables of wrongdoing and responsibility, it fits our moral defenses like a glove.

In this way, the realist interpretation of Strawson's biconditional can appeal to our understanding of excuses to provide a window onto the nature of responsibility.⁹ An analysis of criminal law doctrines of excuse can be a part of this investigation. In this context, it is worth addressing the relationship between excuses and exemptions. The prototypical case of an exemption is a case in which an actor is not responsible for what he did because of quite general impairments of his agency. So, for instance, insanity and immaturity are sometimes described as exemptions. By contrast, excuses are sometimes claimed to be prototypically case-specific in which the agent is otherwise normal and responsible but acted inadvertently or was subject to coercion in a specific situation. Despite the existence of these two different kinds of prototypical cases, we think that it is a mistake to treat exemptions and excuses as disjoint classes. First, while Strawson and others include insanity among the exemptions, the criminal law treats insanity as an excuse. In fact, the criminal law includes all claims to less than full culpability in the single category of excuse. So there is some reason not to assume that exemptions cannot be excuses. Second, the prototypical cases are not exhaustive of the possibilities, as Strawson himself recognized (1962: 79). Strawson's partition is into cases in which the reactive attitudes are generally disabled in regard to a particular agent and cases in which they are selectively disabled due to inadvertence or compulsion. But there are at least three different dimensions on which these paradigm cases can be distinguished: *scope*, *duration*, and the *location* of the obstacle to culpability. Immaturity, for example, or even more temporary conditions, such as depression or even dementia due to dehydration, might undermine responsibility for all sorts of actions during the episode in question, and so have wide scope. In contrast, a particular perceptual deficit, or a compulsive disorder narrowly confined to one area, like kleptomania, might have a relatively narrow scope. Paradigm cases can also be distinguished on the basis of duration. A phobia, for example, might affect one's choices in a narrow area, but be life-long, in contrast to a short spell of dementia caused by dehydration. The third dimension is the location of the obstacle as either within or outside of the agent. Immaturity is an example of the former, and low lighting conditions that prevent one from seeing someone else in need is an example of the latter. All three of these are separable in principle, but in the original paradigm cases, go together. For example, childhood is long-lasting, has a wide scope (though narrowing as one ages), and seems to

⁹Moore describes excuse as the "royal road" to responsibility (1997: 548). Whereas the realist regards our practices of excuse as potential evidence of a response-independent conception of responsibility, a response-dependent conception will understand our practices of excuse as constitutive of responsibility.

be explained by the agent's own capacities. Not realizing one is stepping on another's toes, in contrast, is typically short-lived, narrow in scope, and explainable by something about the particular situation rather than one's capacities. Recognizing that considerations that mitigate culpability can fall in a variety of places along all three dimensions suggests to us that it would be most useful to consider all of the cases as ones involving potential excuses with varying degrees of scope and duration and with varying locations between the agent and the situation. On the proposal that we favor, exemptions are best understood as comparatively global or standing excuses.

The challenge for the realist interpretation of Strawson's thesis is to use the reactive attitudes as evidence to uncover an independent conception of responsibility that can support the reactive attitudes. If we study responsibility by studying excuses, we find that excuses factor into two main kinds on the location dimension. Some excuses reflect compromised psychological capacities of agents. We will conceptualize these as failures of *normative competence*. Insanity is the most familiar excuse of this type. But some excuses reflect no failure of normative competence. Instead, they reflect a lack of normal *situational control*. In such situations, though the agent is normatively competent, factors external to her deprive her of the fair opportunity to avoid wrongdoing. Coercion and duress provide excuses of this type. We think that attention to these two kinds of excuse provides the key to understanding the architecture of responsibility.

2. NORMATIVE COMPETENCE

If someone is to be culpable or responsible for her wrongdoing, then she must be a responsible agent. So we need to distinguish between responsible and non-responsible agents. Our paradigms of responsible agents are normal mature adults with certain sorts of capacities. We do not treat brutes or small children as responsible agents. Brutes and small children both act intentionally, but they act on their strongest desires or, if they exercise deliberation and impulse control, it is primarily instrumental reasoning in the service of fixed aims. By contrast, we suppose, responsible agents must be *normatively competent*. They must 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 respond to 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.¹⁰

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

¹⁰In framing our approach to the internal dimension of responsibility this way, we draw on previous work in the compatibilist tradition that emphasizes normative competence (Wolf 1990, Wallace 1994) and reasons-responsiveness (Wolf 1990, Wallace 1994, Fischer and Ravizza 1998, and Nelkin 2011) and distinguishes cognitive and volitional dimensions of reasons-responsiveness (Wallace 1994, Fischer and Ravizza 1998).

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. It is a condition of our holding them responsible that they possessed the relevant capacities.¹¹

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. 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.

Reasons-responsiveness is clearly a modal notion and admits of degrees; one might be more or less responsive. This raises the question how responsive someone needs to be to be responsible. This is an important and difficult issue, deserving more careful discussion than we can give it here. We make some preliminary remarks here, which we will refine in later sections. We might begin by distinguishing different *grades* of responsiveness. Here, we adapt some ideas from John Fischer and Mark Ravizza in their book *Responsibility and Control* about the responsiveness of the mechanisms on which agents act to our issue about how reasons-responsive the agents themselves are.¹² We propose to specify the degree to which an agent is responsive to reasons in terms of counterfactuals about how she would believe or react in situations in which there was sufficient reason for her to do otherwise.¹³ An agent is more or less responsive to reason depending on how well her judgments about what she ought to do and her choices would track her reasons for action.

We could begin this process by distinguishing two extreme degrees of responsiveness.

¹¹Sidgwick famously objects to Kant's conception of autonomy as conformity to principles of practical reason that this would prevent us from holding criminals responsible and would allow us to recognize only morally upright behavior as responsible (Sidgwick 1907: 511-16). The solution to this problem is for Kant to define autonomy in terms of *capacities* for conformity to principles of practical reason.

¹²The conception of reasons-responsiveness that Fischer and Ravizza defend is mechanism-based, rather than agent-based (1998: 38). By contrast, we favor a version of reasons-responsiveness that is agent-based, rather than mechanism-based, precisely because we think that responsibility and excuse track the agent's capacities, rather than the capacities of her mechanisms. For defense of the agent-based approach, see Nelkin 2011: 64-79 and McKenna 2012.

¹³For present purposes, in specifying an agent's capacities in terms of such counterfactuals, we can remain agnostic about whether capacities or counterfactuals have explanatory priority, in particular, whether capacities ground the counterfactuals or whether the capacities just consist in the truth of such counterfactuals.

- *Strong Responsiveness*: Whenever there is sufficient reason for the agent to act, she recognizes the reason and conforms her behavior to it.
- *Weak Responsiveness*: There is at least one situation in which there is a sufficient reason to act, and the agent recognizes that reason and conforms her behavior to it.

However, it does not seem plausible to model normative competence in terms of either strong or weak responsiveness. Strong responsiveness is too strong for the same reason we gave for focusing on competence, rather than performance. We do not require that people actually act for sufficient reasons to do otherwise; it is the capacities with which they act that matter. The weak-willed are, at least typically, responsible for their poor choices. Moreover, weak responsiveness seems too weak. It treats someone as responsive in the actual situation even if she did not respond in the actual situation and there is only one extreme circumstance in which she would recognize and respond to reasons for action. The Goldilocks standard of responsiveness evidently lies somewhere between these extremes. Of course, there is considerable space between the extremes – the gap between always and once.

We might stake out an intermediate form of responsiveness in something like the following terms.

- *Moderate Responsiveness*: Where there is sufficient reason for the agent to act, she regularly recognizes the reason and conforms her behavior to it.

Moderate responsiveness is deliberately vague; it specifies a range or space of counterfactuals that must be true for the agent to be responsive. Ideally, we would be able to specify a preferred form of moderate responsiveness more precisely. But what is important for present purposes is that reasons-responsiveness is a matter of degree and that the right threshold for responsibility is probably some form of moderate responsiveness.

So far, this conception of responsiveness is coarse-grained in ways that might prove problematic. For one thing, it lumps together cognitive and volitional dimensions of responsiveness. But if they are independent aspects of normative competence, then we may need to assess responsiveness along these two dimensions separately. Moreover, it is at least conceivable that we might require different degrees of responsiveness in cognitive and volitional dimensions of competence. For instance, Fischer and Ravizza distinguish the cognitive and volitional dimensions of reasons-responsiveness in terms of “reasons-receptivity” and “reasons-reactivity” (respectively). Their conception of reasons-responsiveness is *mixed*, because it treats receptivity and reactivity *asymmetrically*. They combine moderate receptivity and weak reactivity (1998: 81-82). We ultimately reject this asymmetry, but it represents a conception of responsiveness worth considering.

Furthermore, this initial formulation of responsiveness assumes that we consider all situations in which there is sufficient reason to act together. But we may find it more informative to partition possibilities into groups, depending on the kinds of reasons at stake and other aspects of the situations in which agents find themselves. For instance, in deciding whether an agent had sufficient volitional capacity to overcome fears that stood in

the way of her performing her duty, we may think it best to restrict our attention to those counterfactuals in which she faced threats or fears of comparable kind or magnitude.

For these reasons, we may need to make our assessments of the degree of an agent's responsiveness more fine-grained in several ways. We address some of these complications below.

3. THE COGNITIVE DIMENSION OF NORMATIVE COMPETENCE

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, as we will see. 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.¹⁴

Here is one place it might be important to distinguish between the demands of moral and criminal responsibility. Presumably, moral responsibility requires the ability to recognize moral norms, including norms that specify moral wrongdoing, whereas criminal responsibility requires the ability to recognize criminal norms, including norms that specify criminal wrongdoing.¹⁵ The cognitive abilities to recognize these two different kinds of norms might be different, with the result that it might be possible to be criminally responsible without being morally responsible and vice versa.

It is common to contrast reason and emotion. This common contrast might lead one to suppose that the cognitive dimension of normative competence is purely cognitive and does not involve emotion or affect. But this conclusion would be misguided. Emotional or affective deficits may block normative competence by compromising cognitive capacity. For instance, lack of empathy may make it impossible or very difficult to recognize actions as injurious and, hence, legally or morally wrong. There is also evidence that congenital damage to the amygdala, which is thought to be the part of the brain responsible for emotional learning and memory, may prevent the formation of normative or, at least, moral concepts. There is emerging research that shows that psychopathy involves both abnormalities in the amygdala and empathy deficits (Blair et al. 2005) Moreover, psychopaths have been thought to have trouble with a psychological test used to discriminate between moral norms and conventional norms.¹⁶ These findings raise questions about whether psychopaths have moral concepts and so whether they have the cognitive capacity to distinguish moral right from wrong. Even if they lack cognitive moral competence, it doesn't follow that they lack the capacity to recognize legal wrongdoing. It is at least possible that some psychopaths might be criminally responsible without being

¹⁴M'Naghten's Case, 10 Cl. & F. 200, 8 Eng. Rep. 718 (1843).

¹⁵There is a debate about whether the cognitive dimension of the insanity test, expressed in M'Naghten's rule, should be formulated in terms of capacities for recognizing criminal or moral wrongdoing. British criminal law has focused on criminal wrongdoing, and American jurisdictions remain divided.

¹⁶See Blair et al. 2005: 57-59. For some skepticism, see Aharoni et al. 2012.

morally responsible.¹⁷ These are complicated issues that deserve fuller examination, but they illustrate ways in which emotion and affect can have a bearing on the cognitive dimension of normative competence. Here, emotional capacities may be *upstream* from normative cognition.

4. THE VOLITIONAL DIMENSION OF NORMATIVE COMPETENCE

But there is more to normative competence than this cognitive capacity. We assume that intentional action is the product of informational states, such as beliefs, and motivational states, such as desires and intentions. Though 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 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 and temptation.

Here, emotion and appetite are *downstream* from cognition and play a separate, volitional or executive role. 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.¹⁸
- Clinical depression that produces systematic weakness of will in the form of listlessness or apathy.
- Acquired or late onset damage to the prefrontal cortex 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.¹⁹

¹⁷We suspect that severe psychopathy impairs, but does not eliminate, reasons-responsiveness and that it may make for a better moral excuse than a criminal excuse. Contrast Fine and Kennett 2004.

¹⁸Mele 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 difficult or impossible (1990). 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.

¹⁹Phineas 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

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 disease or defect" do not include an abnormality manifested only by repeated criminal or otherwise anti-social conduct.²⁰

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.

Recognizing the volitional dimension of normative competence may require revising the rationality or practical reason conceptions of responsibility employed by criminal law theorists such as Michael Moore and Stephen Morse. Strictly speaking, rationality conceptions of normative competence need not reject the volitional dimension of normative competence. There might be more to rationality than correct belief or knowledge. For instance, one might not count as practically rational unless one's appetites and passions are sufficiently under control to enable one to conform one's will to one's normative judgment.

As far as we can tell, Moore is non-committal on this issue and could agree with these claims about the importance of the volitional dimension of normative competence,

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 Damasio 1994.

²⁰American Law Institute, *Model Penal Code* §4.01, emphasis added. The *Model Penal Code* 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. Notice three differences between MPC and M'Naghten: (a) unlike M'Naghten, MPC includes *volitional, as well as cognitive*, capacities in its conception of insanity; (b) whereas M'Naghten makes complete incapacity a condition of insanity, MPC makes *substantial incapacity* a condition of insanity; and (c) whereas M'Naghten requires only capacity for normative recognition for sanity, MPC requires capacity for normative *appreciation*. Here, we focus only on (a), but all three points of contrast between MPC and M'Naghten are potentially significant.

folding them into claims about rational capacities. However, Morse is skeptical about the volitional dimension of normative competence. In part because his skepticism finds echoes in Fischer and Ravizza's treatment of reasons-reactivity, it is worth considering his complaints about the volitional dimension in some detail.

In his essay "Uncontrollable Urges and Irrational People," Morse critically discusses proposals to treat wrongdoers with irresistible impulses as excused for lack of control. He claims, not implausibly, that many with emotional or appetitive disorders are nonetheless responsible, because they retain sufficient capacity for rationality (2002: 1040). In discussing excuses that appeal to uncontrollable urges, he makes clear that his conception of rationality excludes volitional components.

This ... Essay claims that our ambivalence about control problems is caused by a confused understanding of the nature of those problems and argues that control or volitional problems should be abandoned as legal criteria [for excuse] [2002: 1054].

But why should we abandon a volitional dimension to normative competence and control? Morse focuses on the alleged threat posed by irresistible urges and makes several (incompatible) claims about them: (1) we cannot make sense of irresistible urges, (2) we cannot distinguish between genuinely irresistible urges and urges not resisted, (3) there are no irresistible urges, because under sufficient threat of sanction we can resist any strong urge.

Morse focuses on irresistible urges. This is already problematic, because it ignores the varieties of volitional impairment, which include not just irresistible urges but also paralyzing fears, depression, and systematic weakness caused by damage to the prefrontal cortex.

But consider what Morse does say about irresistible urges. He argues against the claim made by the majority in *Kansas v. Crane* that civil detention be limited to those who are dangerous to themselves or others on account of control problems that are the result of mental abnormality.²¹ Morse plausibly claims that mental disease or abnormality, as such, is irrelevant to excuse (2002: 1034, 1040). All that mental abnormality signals is something about the cause of urges; by itself, it does not signify anything about the agent's capacities, and so cannot serve as an excuse (2002: 1040). That is surely right, but the Court in *Crane* did not say that mental abnormality was sufficient for excuse, at most that it was necessary.²² What was critical, the Court claimed, was whether the urges were sufficiently irresistible to present a control problem. A control problem can be understood as a lack of relevant volitional capacities. So the Court is just not making the fallacious argument that Morse rightly criticizes. Demonstrating that abnormality does not imply incapacity does not show that responsibility does not require volitional capacity. So Morse's criticism of the abnormal cause requirement does not support a rationality conception of agency that eschews volitional capacities.

Morse goes on to claim that the idea of irresistible urges is not coherent and that, even if it was, we could not distinguish between irresistible urges and urges not resisted

²¹*Kansas v. Crane*, 534 U.S. 407 (2002).

²²Insofar as the Court is requiring a mental abnormality, perhaps defined on the disease model, we disagree. It is neither a necessary nor sufficient condition for excuse.

(1994: 1601; 2002: 1062). This is the problem of distinguishing between can't and won't. Finally, he asserts that even if we could distinguish between irresistible urges and urges not resisted, we would find that in actual cases the urges in question would be resistible. In discussing whether an addict's cravings are irresistible, Morse argues that they are not because if you hold a gun to the addict's head and tell him that you'll shoot him if he gives in, he can resist (2002: 1057-58, 1070). This is reminiscent of the sort of weak reactivity that Fischer and Ravizza defend and that Kant requires in *The Critique of Practical Reason*.

Suppose that someone says his lust is irresistible when the desired object and opportunity are present. Ask him whether he would not control his passion if, in front of the house where he had this opportunity, a gallows were erected on which he would be hanged immediately after gratifying his lust. We do not have to guess very long what his answer would be [Kant 1788: 30].

But these different complaints about irresistible urges are all resistible.

First, there seems to be no conceptual problem with irresistible urges. We can conceive of paralyzing emotions or irresistible desires, as Mele does, as emotional states or appetites that stand in the way of implementing the verdicts of practical reason that are virtually unconquerable and uncircumventable (1990). Resistibility is a modal notion. There is a question about how unconquerable or uncircumventable impulses must be to be excusing, and there may be evidential or pragmatic problems about identifying desires that are genuinely irresistible. But the concept of irresistible desires does not seem especially problematic.

Second, consider the worry that we cannot reliably distinguish between an inability to overcome and a failure to overcome such obstacles. First of all, this is an evidentiary problem, not a claim about the ingredients of normative competence. Moreover, this evidentiary problem seems no worse than the one for the cognitive dimension of normative competence, which requires us to distinguish between a genuine inability to recognize something as wrong and a failure to form correct normative beliefs or attend to normative information at hand. Making the distinction between can't and won't is a challenge, but not an insurmountable one, in either the cognitive or volitional case. For instance, there are neurophysiological tests for various forms of affective, as well as cognitive, sensitivity, such as electrodermal tests of empathetic responsiveness (Blair et al. 2005: 49-50).

Finally, consider Morse's claim that volitional capacity is easily demonstrated insofar as agents can always resist desires and temptations under sufficient threat. Morse's position here bears comparison with that of Fischer and Ravizza. As we saw earlier, while they defend moderate reasons-receptivity, they require only weak reasons-reactivity.²³ In defense of weak reactivity, Fischer and Ravizza claim that reactivity is "all of a piece" -- if you can conform in some cases, even one case, that shows that you can conform in any case. (1998: 73). Kant and Morse seem to agree. There are two problems here. First, they want to recognize an asymmetry between cognitive and volitional capacities. Yet, if reactivity were "all of a piece," then why not say the same thing about receptivity? If one can

²³It is worth noting that Fischer's and Ravizza's view is *doubly* asymmetric insofar as they require receptivity to at least some *moral* reasons, but require reactivity only to reasons in general, not necessarily moral ones (1998: 79). We reject this sort of asymmetry, as well.

recognize some moral reasons, one can recognize any. Or if one can recognize them under some circumstances, then one can recognize them under any. This would be to accept weak reasons-receptivity, which both Morse and Fischer and Ravizza reject. Second, they are committed to claiming, at least about reasons-reactivity, that one can't have weak responsiveness without having moderate responsiveness. Anyone who can resist an urge in one extreme situation can resist it in others. But we see no reason to accept this psychological stipulation. An agoraphobe might have such a paralyzing fear of public spaces that she would be induced to leave her home only under imminent threat of death. There's no reason to assume that we cannot have weak reactivity without moderate reactivity (cf. McKenna 2001, Watson 2001, Pereboom 2006, and Todd and Tognazzini 2008).

Our own view is that weak reactivity is simply implausible as a general reactivity condition on responsibility. Cases in which a person would only react differently under a threat of imminent death, because of a paralyzing fear or compulsion, for example, seem to be cases in which we should excuse.²⁴ If a desire is really only resistible in this one counterfactual case, then we think that the agent is not responsible, or at least not fully responsible, in the actual case. That doesn't mean that we can't detain him if he is dangerous to himself or others, but it would mean that it would be inappropriate to blame and punish him.

On closer inspection, it seems Morse is really ambivalent between two different kinds of skepticism about the volitional dimension of normative competence and its significance. In some moments, he denies that there is any separate volitional dimension to normative competence, beyond the cognitive dimension. At other times, he recognizes the need for a separate volitional dimension but claims that it is easily satisfied because volitional conformity to what one judges right and wrong is "all of a piece". We hope to have shown that neither form of skepticism is especially promising.

5. SITUATIONAL CONTROL

An important part of an agent's being responsible for wrongdoing that she chose and intended consists in her being a responsible agent. This we have conceptualized in terms of normative competence and analyzed 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. The most familiar kinds of excuse – insanity, immaturity, and uncontrollable urges – all 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. Moreover, 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

²⁴Mele's example of the agoraphobe, who will not leave his house, even for his daughter's wedding, but would leave it if it were on fire, seems coherently described as one in which someone is weakly reactive, but nevertheless, not responsible.

harm to oneself or a loved one if one doesn't assist in some kind of wrongdoing, for instance, driving the getaway car in a robbery. 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 (*NE* 1110a9-12). 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.

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.²⁵ 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 (§2.09).

Whereas the situational aspect of responsibility was recognized by classical writers, such as Aristotle, Hobbes, and Locke, it has been less prominent in more recent philosophical discussions of responsibility. Perhaps because of case law and doctrine involving duress, criminal theorists, such as Moore, and Morse, have clearly recognized the importance of the situational component of responsibility (Moore 1997: 554, 560-61; Morse 1994: 1605, 1617; and Morse 2002: 1058). They explain the rationale for this situational component and the associated excuse of duress in terms of the *fair opportunity to avoid wrongdoing*. The idea is that normatively competent agents, through no fault of their own, due to external threat or hard choice may lack the fair opportunity to avoid wrongdoing. In normal cases, this opportunity may just blend into the background, taken for granted. But in cases of duress it is absent. This not only explains why duress should be

²⁵Exactly when duress justifies and when it excuses is an interesting and difficult question. Much will depend on how the necessity and balance of evils doctrines are understood. Suppose A threatens to rape B's loved one if B doesn't kill C, who is innocent. On one interpretation, this case fails the balance of evils test (murder is worse than rape), so it tends to excuse, rather than justify. But this may be less clear if the balance of evils test is performed using a moral balance employing agent-centered prerogatives. We can't get a clear handle on the difference between duress justifications and duress excuses until we fix the moral conception employed in the balance of evils test. But we think that it is safe to assume that however exactly the lines are drawn on these issues about interpreting the balance of evils test there will be some duress excuses. That is especially plausible, we think, when we recognize that duress and excuse can be scalar. That is sufficient to justify our architectural assumption that there should be a separate wing for situational control, whether or not that wing is densely populated.

excusing but also alerts us to the importance of this opportunity in the normal case, where duress is absent.

6. TWO MODELS OF NORMATIVE COMPETENCE AND SITUATIONAL CONTROL

We think that this emerging picture of the architecture of responsibility in which normative competence and situational control are the two main elements of responsibility is quite attractive. Others have thought so too. For example, in “Negligence, Mens Rea, and Criminal Responsibility” H.L.A Hart seems to accept such a view.

What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for abstaining from what it [the law] forbids, and a fair opportunity to exercise these capacities [Hart 1961: 152].

And Moore endorses this idea.

Hart thus subdivides the ability presupposed by his sense of 'could' into two components. One relates to the equipment of the actor: does he have sufficient choosing capacity to be responsible? The other relates to the situation in which the actor finds himself: does that situation present him with a fair chance to use his capacities for choice so as to give effect to his decision [Moore 1997: 554]?

We see two different conceptions of how these two factors relate to responsibility.

On one conception, normative competence and situational control are individually necessary and jointly sufficient but independent factors in responsibility. On this conception, there is an appropriate degree of competence and an appropriate degree of situational control that can be fixed independently of each other and which are both necessary for responsibility, such that falling short in either dimension is excusing. On this picture, we assess an agent in each area separately. We figure out whether she had the relevant capacities (e.g., were they “normal” or “sufficient”), and then we figure out whether she had the fair opportunity to exercise them.

This has been the conception of the architecture of responsibility that we have articulated so far. However, an alternative conception of normative competence and situational control is possible that treats them as individually necessary and jointly sufficient but at least sometimes interacting. On this picture, how much and what sort of capacities one needs can vary according to situational features. So, for example, there might be situations in which the wrongdoing in question was especially clear, such as a murder or an assault, and in which there was no significant provocation, duress, or other hard choice. We might think that culpable wrongdoing, in such cases, requires less in the form of cognitive or volitional capacities than in cases in which the normative issues are less clear or in which there is substantial provocation or duress. Or hold constant the wrongdoing in question and compare the interaction of situational factors and competence in different individuals. It's plausible to suppose that normative competence requires an ability to make one's own normative judgments and hold to them despite temptation, distraction, and peer pressure. It's also plausible to suppose that adolescents have less independence of judgment and ability to resist peer pressure than their adult counterparts. But then we might be more excusing of adolescent wrongdoing committed in groups than

of comparable adult behavior committed in groups. If so, differential competence may explain why the same level of situational pressure may be excusing for some and not for others (Brink 2004). But then we might prefer a conception of normative competence and situational control in which they are potentially interacting, rather than independent, dimensions of responsibility. Such a conception would also imply that the requisite levels of normative competence and situational control are not invariant, but rather context-dependent.

7. RESPONSIBILITY AS THE FAIR OPPORTUNITY TO AVOID WRONGDOING

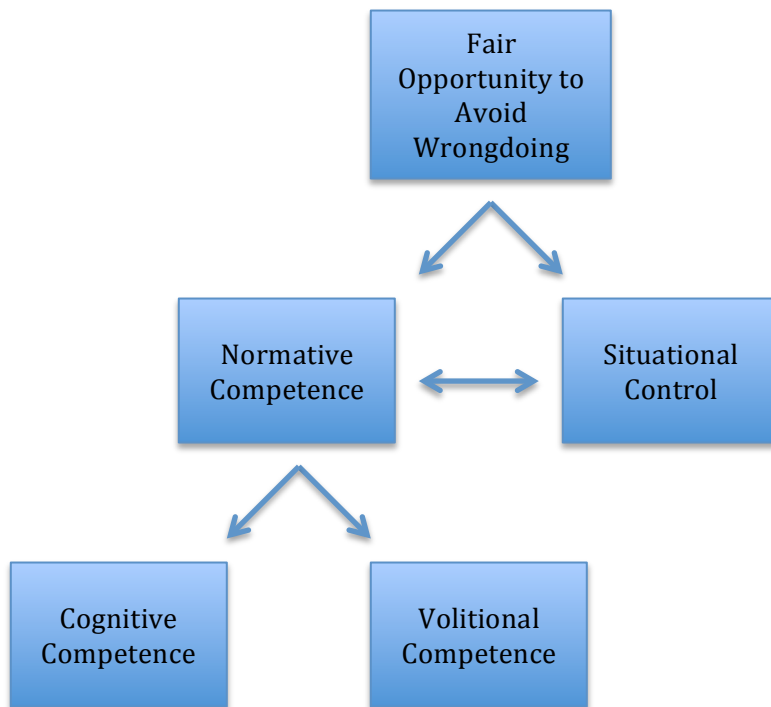
So far, the conception of responsibility emerging from our discussion is a two-factor model twice over. We factor responsibility into normative competence and situational control, and we factor normative competence into partially independent cognitive and volitional capacities, which we treat as equally important. 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* (the concept Fischer and Ravizza associate with reasons-responsiveness). 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.

But this suggests that the umbrella concept should perhaps be *fairness*, in particular, the fair opportunity to avoid wrongdoing (the concept that Moore and Morse associate with the situational component), 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.

A related proposal would be to treat the umbrella concept as the fair opportunity to avoid *blame*. For we might also say that fairness requires that we not blame those who lacked normative competence or situational control. While the opportunity to avoid blame is no doubt a by-product of the fair opportunity to avoid wrongdoing, we think that it is the latter concept that should be primary in our understanding of responsibility. For one thing, if we understand responsibility (as accountability) as the condition that justifies blame, then appeal to the fair opportunity to avoid blame threatens to introduce circularity into our account. For we would then be explaining what makes actions potentially blameworthy in terms of the fair opportunity to avoid blame. That seems like too small an explanatory circle. Moreover, making the primary focus on blame seems potentially narcissistic -- too focused on the agent's moral ledger. It may be that one can avoid blame if one has the fair opportunity to avoid wrongdoing, but being responsible does not seem to be primarily about the agent avoiding blame. Rather, the ability to avoid blame seems consequential on responsibility. We can explain this, without focusing unduly on the responses of others, if we treat the fair opportunity to avoid wrongdoing as the primary concept.

If we treat the fair opportunity to avoid wrongdoing as the key to responsibility, we get the following picture of 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. While tort law recognizes some strict liability offenses, it is significant that the penalties for such torts are civil monetary damages. Strict liability is extremely problematic within the criminal law, where guilt results in blame and imprisonment. Indeed, the main strict liability crime is statutory rape (consensual intercourse by an adult male with an underage female), where a reasonable belief that the female was not a minor is not exculpatory. But this conception of statutory rape is anomalous. Statutory rape is not treated as a strict liability crime in all jurisdictions, and where it is so treated, it is widely viewed as morally problematic. Indeed, the *Model Penal Code* condemns criminal offenses for which conviction does not require culpability (§2.02).²⁶ Why exactly are strict liability crimes so problematic?

In “Legal Responsibility and Excuses” H.L.A. Hart suggests that the criminal law conditions liability on culpability out of respect for “the efficacy of the individual’s informed and considered choice in determining the future” (1957: 46). A corollary of this concern with individual autonomy is the demand for the fair opportunity to avoid wrongdoing. This principle is at work in support of the fundamental legal principle of *legality*. Legality is the doctrine that there should be no punishment in the absence of public notice of a legal requirement. The principle of legality is usually defended as part of fair notice. Ex post facto or retroactive criminal law would be unfair, because it would punish those for failing to conform to behavioral expectations of which they had not been apprised in advance. Ex post facto law thus threatens individual autonomy and its demand

²⁶For a brief discussion of the anomalous position of strict liability crimes within the criminal law, see Dressler 2009: ch. 11.

of fair opportunity to avoid wrongdoing. But a similar rationale is at work against strict liability crimes. Just as it would be unfair to convict actors for failing to conform to standards that had not been promulgated in advance, so too it would be unfair to convict actors for failing to conform to standards (promulgated in advance) that they did everything within their power to obey. Conviction without culpability denies the fair opportunity to avoid wrongdoing.

Some have suggested that fairness requires normative competence as a condition of responsibility.²⁷ Others have suggested that the fair opportunity to avoid wrongdoing requires situational control. We put these ideas together and conclude that the guiding principle underlying our conception of responsibility is the fair opportunity to avoid wrongdoing.

Because the fair opportunity to avoid wrongdoing is the central value in our conception of responsibility, it is perhaps worth addressing skepticism about the fair opportunity to avoid wrongdoing. There is a familiar kind of skepticism about responsibility that appeals to incompatibilism and determinism. One form of skepticism alleges that responsibility presupposes the Principle of Alternative Possibilities (PAP), requiring that one could have done other than one did, which determinism shows to be false. Our conception of responsibility is not immune to these sorts of skeptical threats, because we understand responsibility in terms of the fair opportunity to avoid wrongdoing and we understand the fair opportunity to avoid wrongdoing to require alternative possibilities. To have the fair opportunity to avoid wrongdoing, it must be true that when one commits wrong, one could have done otherwise. There are two kinds of challenge to PAP. First, PAP may seem to require indeterminism. Second, alternative possibilities may seem unnecessary for responsibility because of Frankfurt-style counterexamples in which an agent chooses to act in a normal and responsible way and in which someone else (the

²⁷See Wallace 1994: 109, 116, 161-66. The substance of our conception of responsibility agrees with Wallace on many points, including the idea that responsibility requires normative competence, that normative competence has both cognitive and volitional dimensions, and that normative requirement is a condition of the fairness of blame and punishment. But there are important differences. Wallace's account of excusing and exempting conditions (chs. 5-6) fails at many points to distinguish properly between justification and excuse. Also, like Fischer and Ravizza and other philosophers, Wallace tends to focus on the kind of normative competence that fairness requires for responsibility, thus ignoring or at least underestimating the importance of the kind of situational control that fairness also requires for responsibility. Finally, and perhaps most importantly, Wallace insists on adopting a response-dependent interpretation of this conception of responsibility insofar as he claims that we cannot give an account of the conditions under which an agent is responsible independently of our reactive attitudes and practices of excuse (1994: 19). But while the reactive attitudes and our practices of excuse may be *evidence* about the conditions of responsibility, on our conception, they point to a conception of responsibility grounded in the fair opportunity to avoid wrongdoing, which itself requires normative competence and situational control. On this conception, it is possible to give a response-independent account of the conditions under which the reactive attitudes are justified.

counterfactual intervener) stands ready to intervene to cause her to act that way in case she should waver (Frankfurt 1969).

These are familiar philosophical issues about responsibility that are quite complicated, and this is not the place to defend compatibilism in any detail. But we can and should indicate the lines along which we would like to defend our commitments. Our response is to understand the “could have done otherwise” clause in PAP as consistent with both the existence of typical Frankfurt scenarios and with determinism. This requires a compatibilist understanding of the ability to do otherwise. A variety of such compatibilist accounts have recently been defended.²⁸ The account we favor is one according to which one has the general ability to do otherwise, manifested in various cognitive and volitional capacities and opportunities, and, at the time of the action, nothing impairs these capacities and opportunities in a relevant way. Such a condition can be satisfied in Frankfurt-style cases, and, we claim, in deterministic settings.²⁹ There is, to be sure, another sense of the ability to do otherwise, in which the agents in Frankfurt-style cases lack such an ability, but this is not the sense we believe is required by the fair opportunity to avoid wrongdoing. The legitimate threats to the opportunity to avoid wrongdoing and, hence, responsibility come not from the fact that our actions are caused but rather from specific impairments of cognitive and volitional capacities and specific kinds of external interference with our ability to conform to the relevant norms.

8. PARTIAL RESPONSIBILITY AND EXCUSE

An important feature of our conception of responsibility is that the components of normative competence and situational control are *scalar* concepts that admit of degree. This implies that responsibility is itself scalar and that *partial* responsibility is a real phenomenon.³⁰ Insanity and immaturity are clearly scalar concepts. But so are the ideas of paralyzing emotions, irresistible urges, and a disabling depression. So too is the hardness of a choice, posed by coercion or natural forces, and the question of whether it is more than a reasonable person could resist. Partial responsibility is reasonably well recognized within our assessments of moral responsibility. But the recognition of partial responsibility suggests a basis for criminal jurisprudence reforms.

Criminal law trials are *bifurcated* into two phases -- the *guilt phase* and, if guilty, the *sentencing phase*. Verdicts at the guilt phase are generally *bivalent* -- guilty or not guilty. There are various possible reasons for a finding of not guilty. The prosecution may fail to prove some element of *actus reus* or *mens rea* beyond a reasonable doubt, or the defense

²⁸For example, there have been several attempts to reconcile PAP with determinism by understanding the ability to do otherwise in terms of the having of a disposition. See, for example, Smith 2003, Vihvelin 2004, Fara 2008. For criticism, see Clarke 2009.

²⁹The skeptic might reply that a determined past prevents one from exercising one’s capacities. But, as Nelkin has argued (2011: 64-79), the fact that the past is deterministic does not make it interference or prevention in the relevant sense any more than is the counterfactual intervener in a Frankfurt scenario. See Wolf (1990: 94-116) for a different kind of defense of the conclusion.

³⁰One recent discussion of the scalar character of reasons-responsiveness is Coates and Swenson 2012. For a defense of the scalar character of immaturity and the resulting partial responsibility appropriate for juvenile justice, see Brink 2004.

may provide adequate evidence of justification, by demonstrating necessity, or excuse, by demonstrating duress or insanity. But the factors determining both justification and excuse would appear to be scalar. One might conclude that in an ideal world guilt verdicts should be *multivalent* reflecting the degree of justification or excuse the defendant possessed. Insofar as our bivalent system recognizes only perfect necessity as a justification and complete or substantial lack of normative competence as an excuse, its conception of guilt and punishment is over-inclusive and punishment is not proportional to just deserts. Perhaps there are legitimate pragmatic reasons for opposing an infinitely multivalent system. But even a *trivalent* system that recognized *partial guilt*, as well as guilty and not-guilty verdicts, would provide more retributive justice.

Justifications do not bear on responsibility, as excuses do. So we focus on partial excuse, rather than partial justification. While some European criminal justice systems recognize partial excuse, American criminal law does not. Instead, we recognize only full excuses, relegating considerations that are relevant to excuse but fall short of full excuse to the sentencing phase of a trial.³¹ Some commentators have called for the adoption of a trivalent system for the guilt phase in which the accused can be found guilty, not guilty, or guilty but partially responsible.³²

A trivalent system would be superior to our current bivalent one. Supplementing a bivalent system with sentence mitigation is an imperfect proxy for recognizing partial excuses. Though one can mitigate sentences to respond to various non-culpability factors (e.g. mercy or forgiveness), mitigation is a poor response to reduced culpability. First, mitigation is not always possible where there are mandatory sentencing rules. Second, mitigation, where possible, is largely discretionary. But it shouldn't be discretionary whether to hear and evaluate considerations that would tend to excuse. If complete incompetence is relevant to the guilt phase and a full excuse, then partial competence is also relevant to the guilt phase and a partial excuse, because it appeals to the very same factors that a full excuse appeals to, only to a reduced degree. In effect, the claim is that excuses, whether full or partial, belong to the guilt phase of a trial and that sentence mitigation should be confined to non-culpability considerations (or, perhaps, in a trivalent system, to those culpability considerations that fall below the threshold necessary for partial excuse).

Our conception of the architecture of responsibility recognizes the scalar nature of the components of responsibility and so justifies findings of partial responsibility. If, as the positive retributivist believes, punishment and the reactive attitudes should be proportional to culpable wrongdoing, then just deserts require that we recognize partial excuses. Otherwise, we punish more than just deserts. As we have seen, not all retributivists are of this sort. Yet all retributivists, positive and negative, believe that punishment and the reactive attitudes should not be *out of* proportion to culpable wrongdoing. If this is correct, then justice requires that we recognize partial excuses. Presumably, our moral assessments and practices of moral blame can and do already

³¹The 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.

³²See Morse 2003. Morse's position is especially interesting, because this defense of partial excuse is a reversal of his previous skepticism about partial excuse (Morse 1984).

reflect recognition of partial excuses, albeit imperfectly. The conception of responsibility we have sketched suggests that we should recognize a doctrine of generic partial excuse in the criminal law as well and abandon that doctrine only if and when it proves administratively unfeasible or to have worse retributive outcomes overall.

BIBLIOGRAPHY

- Aharoni, E., Sinnott-Armstrong, W., and Kiehler, K. (2012) "Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction" *Journal of Abnormal Psychology* 121: 484-97.
- American Law Institute. (1981) *Model Penal Code*, ed. M. Dubber (New York: Foundation Press, 2002).
- Aristotle. (*NE*) *Nicomachean Ethics*, trs. T. Irwin (Indianapolis: Hackett, 1985).
- Bedau, H. and Kelly, E. (2010) "Punishment" *The Stanford Encyclopedia of Philosophy* <<http://plato.stanford.edu/archives/spr2010/entries/punishment>>.
- Blair, J., Mitchell, D., and Blair, K. (2005) *The Psychopath: Emotion and the Brain* (Oxford: Blackwell).
- Brink, D. (2004) "Immaturity, Normative Competence, and Juvenile Transfer: How (Not) to Punish Minors for Major Crimes" *Texas Law Review* 82: 1555-85.
- Clarke, R. (2009) "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism" *Mind* 118: 323-51.
- Coates, D.J. and Swenson, P. (2012) "Reasons-Responsiveness and Degrees of Responsibility" *Philosophical Studies*.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam).
- Dressler, J. (2009) *Understanding Criminal Law*, 5th ed. (Boston: LexisNexis).
- Duff, A. (2008) "Legal Punishment" *The Stanford Encyclopedia of Philosophy* <<http://plato.stanford.edu/archives/fall2008/entries/legal-punishment>>.
- Fara, M. (2008) "Masked Abilities and Compatibilism," *Mind* 117: 843-65.
- Fine, C. and Kennette, J. (2004) "Mental Impairment, Moral Understanding, and Criminal Responsibility" *International Journal of Law and Psychiatry* 27: 425-43.
- Fischer, J. and Ravizza, M. (1998) *Responsibility and Control* (New York: Cambridge University Press).
- Frankfurt, H. (1969) "Alternate Possibilities and Moral Responsibility" reprinted in Frankfurt 1988.
- Frankfurt, H. (1988) *The Importance of What We Care About* (New York: Cambridge University Press).
- Hart, H.L.A. (1957) "Legal Responsibility and Excuses" reprinted in Hart 1968.
- Hart, H.L.A. (1961) "Negligence, Mens Rea, and Criminal Responsibility" reprinted in Hart 1968.
- Hart, H.L.A. (1968) *Punishment and Responsibility* (Oxford: Clarendon Press).
- Kant, I. (1788) *Critique of Practical Reason*, trs. L.W. Beck (Indianapolis: Bobbs-Merrill, 1956) (Prussian Academy pagination).
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language* (Cambridge: Harvard University Press).
- McKenna, M. (2001) "Review of Fischer and Ravizza, *Responsibility and Control*" *Journal of Philosophy* 98: 93-100.
- McKenna, M. (2012) "Reasons-Responsiveness, Agents, and Mechanisms" (this volume).

- Mele, A. (1990) "Irresistible Desires" *Nous* 24: 455-72.
- Moore, M. (1997) *Placing Blame* (Oxford: Clarendon Press).
- Morse, S. (1984) "Undiminished Confusion in Diminished Capacity" *The Journal of Criminal Law and Criminology* 75: 1-55.
- Morse, S. (1994) "Culpability and Control" *University of Pennsylvania Law Review* 142: 1587-1660.
- Morse, S. (2002) "Uncontrollable Urges and Irrational People" *Virginia Law Review* 88: 1025-78.
- Morse, S. (2003) "Diminished Rationality, Diminished Responsibility" *Ohio State Journal of Criminal Law* 1: 289-308.
- Nelkin, D. (2011) *Making Sense of Freedom and Responsibility* (Oxford: Clarendon Press).
- Pereboom, D. (2006) "Reasons-Responsiveness, Alternative Possibilities, and Manipulation Arguments against Compatibilism" *Philosophical Books* 47: 198-212.
- Pereboom, D. (2012) "Freedom and Punishment" in *The Future of Punishment*, ed. T. Nadelhoffer (Oxford: Clarendon Press).
- Scanlon, T.M. (2008) *Moral Dimensions: Permissibility, Meaning, and Blame* (Cambridge: Harvard University Press).
- Sidgwick, H. (1907) *The Methods of Ethics*, 7th ed. (London: Macmillan).
- Smith, M. (2003) "Rational Capacities, or How to Distinguish Recklessness, Weakness, and Compulsion" in *Weakness of Will and Practical Irrationality*, ed. S. Stroud and C. Tappolet (Oxford: Clarendon Press).
- Strawson, P.F. (1962) "Freedom and Resentment" reprinted in Watson 1982.
- Wallace, R. J. (1994) *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press).
- Todd, P. and Tognazzini, N. (2008) "A Problem for Guidance Control" *Philosophical Quarterly* 58: 685-92.
- Vihvelin, K. (2004) "Free Will Demystified: A Dispositional Account," *Philosophical Topics* 32, 427-50.
- Watson, G., ed. (1982) *Free Will* (New York: Oxford University Press).
- Watson, G. (1987) "Responsibility and the Limits of Evil" reprinted in Watson 2004.
- Watson, G. (1996) "Two Faces of Responsibility" reprinted in Watson 2004.
- Watson, G. (2001) "Reasons and Responsibility" reprinted in Watson 2004.
- Watson, G. (2004) *Agency and Answerability* (New York: Oxford University Press).
- Watson, G. (2011) "The Trouble with Psychopaths" in *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (New York: Oxford University Press).
- Williams, B. (1976) "Moral Luck" reprinted in Williams 1981.
- Williams, B. (1981) *Moral Luck* (Cambridge: Cambridge University Press).
- Wolf, S. (1990) *Freedom Within Reason* (New York: Oxford University Press).