accordance with their native morality.[12] They do not *accept* that killing to save honor is wrong.

In the case of honor killings, it is fairly clear that failure to accept or adopt a moral or legal norm does not deprive someone of responsibility. Honor killings are cases of moral intransigence, not of moral inability. They are no different from typical homicides. Neither killer believes they are truly culpable. However, this lack of recognition is not usually thought to be mitigating. By contrast, it is often thought to make the offender *more* culpable. For their action is not the result of a mistake or an accident, it is the result of adopting values that we find despicable. It is their beliefs about the moral value of Jews, in particular, that make Nazis so despised. It is the belief that women are chattel and have no value other than to serve men and bear male children that makes many of the practices in the Middle East so horrific. The inferential link we require for responsibility, therefore, must be to the current values of the person in question *or* to values that we judge that she *ought* to have had under the circumstances.

If we assume that culture does not inhibit people's ability to evaluate or change their values—I shall argue for this shortly—then what is at issue with honor killers is whether they *hold on to* their values or whether they do not have access to relevant values and information that would enable them to change them in the right way. Wolf suggests that if one is brought up in a particular moral milieu, which allows, perhaps encourages, certain wrongs, one is not really in a position to accurately evaluate or change those values. It remains obscure, however, why *we* have the capacity to change our wrongheaded values. My hunch is that Wolf thinks that there is a pretty straightforward inferential route from our current moral values to future, improved, ones. For instance, our indifference to the death and suffering of nonhuman animals, particularly the ones that we eat, is quite plausibly culpable. We have the capacity to arrive at valuing the life and well-being of nonhuman animals because of values and beliefs that we either possess or that are readily available in our environment. We believe that if an action or policy creates unnecessary or avoidable suffering, then we have a prima facie reason not to perform or institute it. We also know—or if we do not actually know, we could easily come to know—that factory farming creates a great amount of suffering, and that we do not need to consume as much meat as we do for proper nutrition. From those beliefs, there is a relatively straightforward inferential route to the belief that factory farming is wrong and that we ought to oppose policies that permit it. As we shall see, this line

[12] This is not to say that many people living in those countries do not regard honor killings with horror, and would never engage in, nor condone such acts.

of reasoning shows not only that *we* are responsible, but also that there usually is little question of someone's culture inhibiting her responsibility. If *we* are responsible for our wrongdoings, so are the slaveholders of Ancient Greece and the male chauvinists of our father's generation. If they are not, neither are we.

Michele Moody-Adams (1994) has argued that people in cultures that permit, or encourage, practices that we condemn, such as slavery, are typically exercising so-called affected ignorance. That is, they chose not to question, seek information or otherwise know about these practices of wrongdoing.[13] I suspect that there is another form of affected ignorance that derives from the degree of difficulty involved in endorsing values that significantly diverge from the culture at large, not to mention advocating a societal change of standards. It requires some imagination to envisage a different moral order. It is, perhaps, an affected lack of imagination. Wolf maintains that the ancient Greeks cannot be held responsible for their attitudes towards slavery. The question, however, is whether there is a not too onerous inferential route from beliefs and values that the ancients possessed that leads to the recognition that slavery is morally wrong.

At the time of its practice, slavery was widely regarded as a terrible fate. When Andromache bewails the death of Hector in *The Illiad*, she decries the fate of the citizens of Troy: "all who will soon be carried off in the hollow ships and I with them—And you, my child, will follow me to labor, somewhere, at harsh, degrading, work, slaving under some heartless master's eye" (Book 24, line 860 ff.).[14] In Xenophon's *Symposium*, Antisthenes suggests that enslaving others is a crime: "Want prompts a thousand crimes, you must admit. Why do men steal? why break burglariously into houses? why hale men and women captive and make slaves of them? Is it not from want?" (Xenophon 2008a: §27) In *Hellenica*, Xenophon not only talks about the great lengths that people will go to, to avoid being "reduced to" slavery, but also recounts of the Spartan general Callicratidas who refused to enslave the Methymnaeans because they were fellow Hellenes (Xenophon 2008b). In the *Politics*, Aristotle refers to people who "affirm that the rule of a master over slaves is contrary to nature, and that the distinction between slave and freeman exists by law

[13] Another feature of affected ignorance is the rephrasing of wrongs in relatively inoffensive sounding language. For instance, the message on the Rwandan radio encouraging the Hutus to kill the Tutsis was "cut down the tall trees" (Dallaire 2004).

[14] The sentiment appears to have been typical in all times where slavery was a predictable result of conquest. Thus, in *Beowulf* "[a] Geat woman too sang out in grief; with hair bound up, she unburdened herself of her worst fears, a wild litany of nightmare and lament: her nation invaded, enemies on the rampage, bodies in piles, slavery and abasement." (line 3150 ff.)

only, and not by nature; and being an interference with nature is therefore unjust." (*Politics*, Book I, Part III) In *Rhetorica ad Alexandrum*,[15] the Sophist Alcidamas says "The deity gave liberty to all men and nature created no one a slave" in reference to the Thebans freeing the Messenians, who had been taken slaves by the Spartans (Garlan 1988: 125).

The passages above suggest that it was not unthinkable to the Ancient Greeks that slavery was wrong. Clearly some people thought slavery was unjust. Furthermore, the elements for a realization of the moral wrongness of slavery were there. Every free person wanted to remain free and regarded slavery as degrading and awful (1988). They recognized that slaves were fellow human beings, that they were capable of suffering, that suffering was morally relevant, that they themselves might be in danger of enslavement by others, and so on. All the talk of the baseness and stupidity of slaves that we also find in the extant literature seems designed to protect an affected ignorance of the wrongness of the institution. After all, as long as *you* are not a slave but *others* are your slaves, it is to your advantage to embrace a norm that permits slavery. If we were to make a comparison to current culture, we might point out that we, too, are in the possession of everything we need to know to recognize that factory farming, for instance, is an immoral practice. Far from being *incapable* of recognizing that this is so, we are choosing to ignore it because it is difficult to imagine not eating meat or eating meat much more rarely, to imagine a societal change, to eat differently from everybody else, and so on, not to mention the economic difficulties that would be involved. Despite these difficulties, however, we are hardly *unable* to recognize the wrongness of the practice.

There are other features of cultural value systems that speak against them having the power to deprive agents of responsibility. It is no secret that to talk of *a* value system of a particular culture is something of an idealization. Though there may be agreement about the most serious forms of transgressions, a society is not characterized by complete agreement about moral norms. Furthermore, such values are subject to change. And, as a matter of historical fact, values *do* change. Such change may be a relatively fluid affair, or be more cataclysmic. Now, *people* instantiate or embody norms (Moody-Adams 1994). For change in values to be possible, people must be able to change their norms: evaluate them, adopt them, defend them, relinquish them, overthrow them, and so on. The point is obvious on reflection. *If* it were true that people brought up in societies where slavery was

[15] Written around the same time as *Rhetoric,* it was traditionally attributed to Aristotle, but might have been written by Anaximander.

sanctioned by law and common morality were thereby incapable of thinking it was wrong, *then* we should expect no moral change. But such change *did* happen; not in Ancient Greece, but in Europe and the Americas. Therefore, adoption of seriously wrongheaded norms does not, by itself, deprive a subject of responsibility (or reduce said responsibility).

The fact that slavery *was* abolished suggests two things. First, there is an inferential route from values and beliefs the subject did have or values and beliefs that she was capable of gaining access to (without unreasonable hardship) to holding the belief/adopting the value that slavery was wrong. Second, a person who is capable of embracing one set of norms is capable of evaluating and changing said norms. The capacity for self-evaluation and change is, after all, a ubiquitous feature of our abilities. When I was a child, I believed in God, now I do not. More pertinently, perhaps, I once thought abortion was wrong, now I do not. This says something about my belief and value formation abilities. Possessing skills, tastes, habits, and values do not prevent change. To the contrary, it reveals the capacity to acquire, evaluate, update, and change such skills, tastes, habits, and values. Unless the subject has been exposed to some physiological or psychological insult, if she possesses values, she has the capacity to think about them, consider their worth, and change them if required. Consequently, people from other cultures typically have the capacity to self-evaluate and change their values accordingly. Furthermore, in many, if not most, instances their environment is sufficiently rich to contain values and information sufficient for them to change their values to what we now take to be the right ones. Consequently, they can be held responsible for their wrongdoing because their adoption of wrongheaded values was based on affected ignorance.

None of this shows that *all* cases of cultural differences in value are due to affected ignorance, nor that there are no agents who are *genuinely* incapable of comprehending that some of their values are questionable. But we have seen that affected ignorance characterizes many such cases. And where it does not, it is not the capacity to evaluate or change one's values that is at issue, but whether the person's value and environment would have allowed her to reach values close enough to what we think are the right ones. The slavery example suggests that affected ignorance is the main culprit behind divergent moral values. But to make this argument requires a more thoroughgoing exploration. Suffice it to say that affected ignorance is an important factor behind the holding of intransigent and divergent values.

The usual examples of cultural differences in values are, at any rate, quite different from the prototypical cases of insanity. The insane rarely possess values that differ from ours, nor do they suffer from specifically

moral deficits.[16] They typically suffer from delusions and hallucinations that affect many different areas of their lives, not just their values.[17] These delusions or hallucinations create the disturbance that gives rise to the criminal action. For instance, an insane person might think that he is facing a creature very different from the one he is actually struggling with, like the infanticidal father thinking he's fighting a snake; he might think that the person has designs on his life and take himself to be acting in self-defense, as apparently McNaughtan did; or he could believe he is committing a harm only to avoid a greater future harm, as in the case of Yates. Instead of espousing substantially different values, the insane are typically wrong about nonmoral matters of fact. In other cases, their madness creates lacunas in their moral outlook. A person suffering from command hallucinations might think that she ought to carry out an otherwise prohibited action, for instance kill someone. But even in these cases there is rarely a radical change in their moral compass.

When it comes to insanity, then, determinations of someone's responsibility depend in no small measure on whether their belief formation is subject to undue influences as a result of mental disorder (or physiological insult), to what extent, and how it affects their values. Depression is not typically an excusing condition, for though the subject's thoughts are affected by the depression—her life seems lackluster and meaningless—it is unlikely to affect her in such a way that she becomes unable to tell right from wrong or become incapable of acting in accordance with her values.[18] At the other extreme, someone in the grip of psychosis whose vision of the world has been distorted may kill her child in the mistaken belief that she is preventing future greater harm. The prototype of insanity looks little like the prototype of intransigent values that differ from ours.

[16] This is why I think psychopaths are not insane (Maibom 2008).

[17] One might argue that people from different cultures possess a whole range of false beliefs, which play a distorting role similar to that of delusions or hallucinations, and that therefore the analogy between insanity and culturally induced values holds. Not so. First, people who have looked for such differences in beliefs that would be relevant to the morally divergent views have had difficulties finding them (Brink 1989; Doris and Plakias 2008). Second, by contrast to ordinary beliefs, subjects tend to be deeply convinced by the truth of their delusions although they are often bizarre. A delusion is not justified by the available evidence—it is isolated from other relevant beliefs—and is often held despite overwhelming reasons not to believe it. And hallucinations involve as-if perceptions, which is not characteristic of false beliefs generally. Third, insofar as we are all subject to culturally induced beliefs, that are similar to delusions, either we are as little responsible for holding values derived from them as are people from different times or cultures, or we are all responsible.

[18] Or rather, we may excuse her for small failures that seem to flow from her anhedonic state, but should she kill someone we will be skeptical that "the depression made her do it."

4. CONCLUSION

If there is an inferential route from someone's beliefs, values, etc. to knowledge that what she is doing is wrong, and it was not too onerous to follow, then we can hold her responsible. Moreover, if she does not know what she is doing is wrong and she does not possess beliefs or values from which there is an inferential route to such knowledge, but she could acquire such knowledge, beliefs, or values without too much hardship given her environment and her mental condition, she can also be held responsible. This is, I have argued, the best interpretation of the epistemic condition on responsibility, at least when it comes to moral intransigence. Choosing to ignore that what one does is wrong *or* simply disagreeing with one's community about what is morally right or wrong cannot be held up as an excuse.

Wolf is right to point out that *if* a person is unable to objectively evaluate or appropriately change her values, *then* she should be excused *ceteris paribus*. She cannot be held responsible for performing actions that she was unable to know were wrong. However, culture does not, in general, inhibit moral change, nor do divergent upbringings. The very fact that we possess values suggests that we are capable of changing them, barring madness or brain damage. Whether a person can change her values so that they are more in accord with what we now believe are the right ones depends, of course, on her values and beliefs and the information that her environment affords. I argued that ancient Greek slaveholders can be held responsible, and I see little reason to think male chauvinists of our father's generation cannot also be blamed. It may turn out that most cases of culturally divergent values are due not to *inability*, but to *inexpediency*. Unless we simply assume that one can only be held responsible if one thinks of one's action as wrong, we should not suppose that people we usually judge to be the most evil—e.g. perpetrators of genocides, slaveholders, and pedophile rapists—are the least responsible. Such a view reduces all culpability to moral incontinence. But as we have seen, most who do wrong do not take themselves to do so. These wrongs are not typically perpetrated by individuals who are *unable* to see the error of their ways, but by individuals who are *unwilling* or are *neglecting* to do so; they engage in affected ignorance. In the case of cultural differences, there is often an inferential route from values and beliefs that are held by the person to what we take to be the right values. For instance, there is little question that the Hutu killers were able to recognize the humanity of their Tutsi neighbors, that they recognized the prohibition on killing, etc., etc. It was, however, expedient to ignore this, so those who were not coerced into engaging in the genocide chose to ignore the wrongness of their actions. Consequently, they did not think of their actions as wrong. Yet, we can surely hold them responsible.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders,* 4th edn. Text Revision *(DSM-IV)*. (Washington DC: American Psychiatric Association).

Aristotle (1984). "The Politics." Trans. Jowett. In: *The Complete Works of Aristotle*. ed. J. Barnes. (Princeton, NJ: Princeton University Press).

Baumeister, R. (1997). *Evil: Inside Human Violence and Cruelty*. (New York: Henry Holt).

——Stillwell, A., and Wotman, S. (1990). "Victims and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger." *Journal of Personality and Social Psychology* 59, 994–1005.

Beck-Sander, A., Birchwood, M., and Chadwick, P. (1997). "Acting on command hallucinations: A cognitive approach." *British Journal of Clinical Psychology* 36, 139–48.

Beowulf. (2000). Trans. Seamus Heaney. (London: W. W. Norton & Co).

Berkowitz, L. (1978). "Is criminal violence normative behavior? Hostile and instrumental aggression in violent incidents." *Journal of Research in Crime and Delinquency* 15, 148–61.

——and Powers, P. (1979). "Effects of timing and justification of witnessed aggression on the observers' punitiveness." *Journal of Research in Personality* 13, 71–80.

Brandt, R. B. (1959). *Ethical Theory: The Problems of Normative and Critical Ethics*. (Englewood Cliffs, NJ: Prentice-Hall).

Brink, D. (1989). *Moral Realism and the Foundation For Ethics*. (Cambridge: Cambridge Uuniversity Press).

Calvete, E. (2008). "Justification of violence and grandiosity schemas as predictors of antisocial behavior in adolescents." *Journal of Abnormal Child Psychology* 36, 1083–95.

CBC News, June 15, 2010. Father, son plead guilty to Aqsa Parvez murder. <http://www.cbc.ca/news/canada/toronto/story/2010/06/15/parvez-guilty-plea.html>.

Dallaire, R. (2004). *Shake Hands with the Devil: The Failure of Humanity in Rwanda*. (Toronto: Vintage Canada).

Doris, J. and Plakias, A. (2008). "How to argue about disagreement: Evaluative diversity and moral realism." In W. Sinnott-Armstrong (ed.), *Moral Psychology, vol. 2. The Cognitive Science of Morality: Intuition and Diversity*. (Cambridge, MA: MIT Press), 303–31.

Faraci, D. and Shoemaker, D. (2010). "Insanity, Deep Selves, and Moral Responsibility: The Case of JoJo." *Review of Philosophy & Psychology* 1, 319–32.

Frankfurt, H. (2003). "Freedom of the will and the concept of a person." In G. Watson (ed.), *Free Will,* 2nd edn. (New York: Oxford University Press), 322–36.

Garlan, Y. (1988). *Slavery in Ancient Greece*. (Cornell: Cornell University Press).

The Iliad. (1990). Trans..Robert Fagles. (Bath: The Bath Press).

Jeffrey, R., Pasewark, R., and Bieber, S. (1988). "Insanity pleas: predicting Not Guilty by Reason of Insanity adjudications." *Bulletin of the American Academy of Psychiatry and Law* 16, 35–9.

Katz, J. (1988). *Seductions of Crime: Moral and Sensual Attractions in Doing Evil*. (New York: Basic Books).

Krakauer, J. (2003). *Under the Banner of Heaven: A Story of Violent Faith*. (New York: Anchor Books).

Larbey, C. (2007). "The secret lives of Kim John and Francis Philip. What do we really know about the Cathedral killers?" *St. Lucia Star*, May 18, 2007.

Luckenbill, D. (1977). "Criminal homicide as a situated transaction." *Social Problems* 25, 176–86.

McNiel, D., Eisner, J., and Binder, R. (2000). "The relationship between command hallucinations and violence." *Psychiatric Services* 51, 1288–92.

Maibom, H. (2008). "The mad, the bad, and the psychopath." *Neuroethics* 1, 167–84.

Maudsley, H. (1898). *Responsibility in Mental Disease.* (New York: D. Appleton and Company).

Moody-Adams, M. (1994). "Culture, responsibility, and affected ignorance." *Ethics* 104, 291–309.

Moran, R. (1981). *Knowing Right From Wrong*. (New York: The Free Press).

MY Lee, T., Chong, S., Chan, Y., Sathyadevan, G. (2004). "Command hallucinations among Asian patients with schizophrenia." *Canadian Journal of Psychiatry* 49, 838–42.

Reznik, L. (1997). *Evil or Ill: Defending the Insanity Defence*. (New York: Routledge).

Rice, M. and Harris, G. (1990). "The predictors of insanity acquittal." *International Journal of Law and Psychiatry* 13, 217–24.

Rogers, J., Bloom, J., and Manson, S. (1984). "Insanity defences: contested or conceded?" *American Journal of Psychiatry* 141, 885–8.

Schwartz, M., O'Leary, S., and Kendziora, K. (1997). "Dating aggression among high school students." *Violence and Victims*, 295–305.

Steadman, H., Keitner, L., Braff, J., and Arranites, T. (1983). "Factors associated with a successful insanity plea." *American Journal of Psychiatry* 140, 401–405.

Stillwell, A., Baumeister, R., and Del Priore, R. (2008). "We're all victims here: Towards a psychology of revenge." *Basic and Applied Social Psychology* 30, 253–63.

Taylor, C. (1976). "Responsibility for self." In A. E. Rorty (ed.) *The Identities of Persons*. (Berkeley, CA: University of California Press), 281–99.

Watson, G. (1975). Free agency. *Journal of Philosophy*, LXII, 205–20.

Williams, B. (1981). "Internal and external reasons." In his *Moral Luck*. (Cambridge: Cambridge University Press), 101–13.

Wolf, S. (2003). "Sanity and the metaphysics of responsibility." In G. Watson (ed.), *Free Will,* 2nd edn. (New York: Oxford University Press), 372–87.

Wolfgang, M. (1958). *Patterns in Criminal Homicide*. (Philadelphia, PA: University of Pennsylvania Press).

Xenophon. (2008a). *Symposium*. Trans. H. G. Dakyns. Project Guttenberg (accessed January 14, 2012).

Xenophon. (2008b). *Hellenica*. Trans. H. G. Dakyns. Project Guttenberg (accessed January 14, 2012).

Zelli, A., Dodge, K., Laird, R., Lochman, J., and Conduct Problems Prevention Research Group. (1999). "The distinction between beliefs legitimizing aggression and deviant processing of social cues: Testing measurement validity and the hypothesis that biased processing mediates the effects of beliefs on aggression." *Journal of Personality and Social Psychology* 77, 150–66.

12

Fairness and the Architecture of Responsibility[1]

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

[1] This 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.

[2] The 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.

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.

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 1994: 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.

[3] Here, 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. 202).

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 are responsible, but it's the responsibility of the targets

[4] Watson 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.

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 Section 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.

[5] Because 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: 66–7). 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.

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 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 nonculpability moral considerations, such as forgiveness or mercy. By contrast, according to

[6] Cf. 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.

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 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 on to 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

[7] Cf. 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.

[8] Precisely 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.

[9] Moore 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.

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 lifelong, 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 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 nonresponsible 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

[10] In 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

[11] Sidgwick 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.

[12] The 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.

[13] For 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.

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.

- *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: 81–2). 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 finds 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

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

[15] There 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.

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 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

[16] See Blair et al. 2005: 57–9. For some skepticism, see Aharoni et al. 2012.

[17] We 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.

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 nonoptimizing 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]

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

[18] 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 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 that serve alcohol circumvents his impulses. Conquerability is mostly a matter of will power, whereas circumventability is mostly a matter of foresight and strategy.

[19] 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 Damasio 1994.

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 noncommittal on this issue and could agree with these claims about the importance of the volitional dimension of normative competence, 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

[20] American 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.

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, but 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.

[21] *Kansas v. Crane*, 534 US 407 (2002).

[22] Insofar 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.

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 (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: 1057–8, 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 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

[23] It 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.

[24] Mele'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.

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 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 (section 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, 560–1, 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 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.

[25] Exactly 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.

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