CULPABILITY AND CONTROL

STEPHEN J. MORSE†

"I couldn't help myself"; "I had no choice"; "I couldn't control myself"; "I was forced to do it." All are common explanations used to support the claim that an agent is not morally or criminally responsible for otherwise culpable conduct. The most common criminal law "control" excuses that instantiate these claims are duress1 and the so-called "volitional" tests for legal insanity.2 More rarely, the agent uses such allegations to attack the prima facie case, claiming that action was lacking.3 Although, properly speaking, the negation of the voluntary act requirement does not raise an excuse, in such cases we also say that the agent was nonculpably out of control. Finally, these explanations are sometimes synonymous with the general conclusion that, for some reason, the agent is not responsible. If it is true that an agent really could not help or control herself and was not responsible for the loss of control, blame and punishment are not justified on any theory of morality

† Ferdinand Wakeman Hubbell Professor of Law, University of Pennsylvania Law School, and Professor of Psychology and Law in Psychiatry, University of Pennsylvania School of Medicine. I would like to thank Simon Blackburn, Richard Double, Herbert Fingarette, Heidi Hurd, Leo Katz, John Monahan, Michael Moore, and Alan Schwartz for their assistance and friendship. The howlers are solely my responsibility.

1 See MODEL PENAL CODE § 2.09(1) (Proposed Official Draft 1962):
   It is an affirmative defense that the actor engaged in the conduct charged to constitute an offense because he was coerced to do so by the use of, or a threat to use, unlawful force against his person or the person of another, that a person of reasonable firmness in his situation would have been unable to resist.

Id.

2 See, e.g., id. § 4.01(1). Section 4.01(1) provides that "[a] person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity . . . to conform his conduct to the requirements of law." See also Parsons v. State, 2 So. 854, 859 (Ala. 1887) (establishing as a criterion for the insanity defense an inability "to choose" between right and wrong).

3 See, e.g., MODEL PENAL CODE § 2.01(2). Section 2.01(2) states:
   The following are not voluntary acts within the meaning of this Section:
   (a) a reflex or convulsion;
   (b) a bodily movement during unconsciousness or sleep;
   (c) conduct during hypnosis or resulting from hypnotic suggestion;
   (d) a bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious or habitual.

Id.

(1587)
and criminal punishment. But what does it mean to say that an agent could not control herself?

This Article attempts to clarify the meaning of and criteria for "control excuses" and their relationship to cases of no action. I shall use philosophy of mind and action, buttressed by findings from the behavioral sciences, to clarify the meaning of excuse based on lack of control. I begin in Part I with a brief homage to Michael Moore, in which I claim that the explanation of human action is still mysterious, despite the herculean efforts of the metaphysicians, but that criminal law theory nevertheless must grapple with the theory of action. Part II considers false starts: popular but inaccurate, misleading, or confusing ways of defining control problems and excuses. I show that these problems are not created by the truth of determinism or universal causation, by the agent's lack of intention, by defects of the will or volition, by irresistible impulses, or by the agent's inability to exercise choice. The Article then turns in Part III to a very brief account of how we humans manage to behave ourselves in the face of ubiquitous antisocial desires. This part provides a list of "self-protective devices" for the purpose of helping to determine what the real problem is when an agent claims to be "out of control." Part IV canvasses possible adequate justifications for control excuses, arguing that most putative control or volitional problems are best understood as rationality defects. In Part V, I conclude by addressing the practical difficulties besetting the assessment and adjudication of control cases.

I. METAPHYSICS, MYSTERY, AND METHODOLOGY: HOMAGE TO MOORE

Human action, as opposed to mere movements of bodies in space, has always perplexed philosophers and students of the psyche. In overly simplified and polarized terms, some contend that action is epiphenomenal, ultimately reducible to biophysical states and events, whereas their opponents, including Michael Moore,\textsuperscript{5} claim that action is not so reducible and supervenes upon those very states and events.\textsuperscript{6} Within the contending camps there are,

\textsuperscript{4} Throughout this Article, I use the term "out-of-control agent" to refer to any person who claims that a control excuse was warranted on a particular occasion.


\textsuperscript{6} On "supervenience" generally, see Jaegwon Kim, Supervenience as a Philosophical Concept, in SUPERVENIENCE AND MIND 131-60 (1993).
predictably, numerous plausible variations. What is a non-philosopher or even a philosopher to think? The best guess is that the correct account of the metaphysics of action is a mystery still, and, alas, shows no sign of being solved anytime soon.

Nevertheless, action theory cannot be avoided by criminal law theorists: the criminal law presupposes precisely the folk psychological account of human action based on desires, beliefs, intentions, and other mental states that reductionist theories reject. It would be possible, I suppose, for a crime control system to abandon viewing individuals as moral agents who act for reasons and who thereby may be blamed and praised, punished and rewarded, as intrinsically appropriate responses to the moral qualities of their conduct. Instead, the criminal law might treat persons as part of the biophysical flotsam and jetsam of the universe and respond solely on the basis of the type and degree of dangerousness people threaten, without regard to moral responsibility. In such a world, practices such as blame and punishment would have primarily instrumental value, useful for controlling the harm-producing behavior of homo sapiens, much like antibiotics are useful for controlling the dangerous propensities of bacteria. In the Anglo-American world, however, virtually no criminal law theorist or lawyer adheres to the purely consequentialist dystopia just sketched. Nearly all are either pure retributivists, like Moore, or mixed theorists for whom just deserts are a necessary, albeit not sufficient, justification for criminal liability and punishment.

---

9 But see BARBARA WOOTTON, CRIME AND THE CRIMINAL LAW: REFLECTIONS OF A MAGISTRATE AND SOCIAL SCIENTIST 44-51 (2d ed. 1981) (defending the extension of strict liability in the criminal law and relegating questions about mental states to the dispositional phase of criminal proceedings).
11 See, e.g., NORVAL MORRIS, MADNESS AND THE CRIMINAL LAW 148-52, 161-68 (1982) (describing the “limiting retributivist” as believing that “more subtle, finely tuned moral distinctions of desert relevant to punishment can and should be made” and using “utilitarian values” to do so). For an interesting thought experiment suggesting that many mixed theorists are actually pure retributivists, see Moore, Moral Worth of Retribution, supra note 10, at 184-85.
either case, the necessity for just deserts presupposes a view of persons as potentially morally responsible, that is, as rational, uncompelled agents, rather than as merely bodies moving in space.

Given the criminal law's reliance on some theory of human action, it would be utterly astonishing if the philosophy of action and its implications did not clarify many of the central puzzles of structure and doctrine. And as Act and Crime demonstrates, the voluntary act doctrine, the actus reus requirement, and the criteria for double jeopardy can be immensely enriched by just the sort of abstract, conceptual analysis Michael Moore provides. I believe this Article demonstrates that this is also true of control problems and excuses. To reject the relevance of action theory, to fail to wrestle with the mystery of action, is both implicitly to do it anyhow\(^\text{12}\) and explicitly to abandon the quest for deeper understanding.

II. COMMON BUT FALSE STARTS, OR WHY I WISH PEOPLE WOULD STOP TALKING CERTAIN WAYS

A. The Lure of Mechanism and Metaphor

Suppose you are sitting on a table with your legs dangling over the edge. Assume that you are neurologically intact and that you strongly prefer not to raise your legs because there is a good, generalizable reason to remain still. Here are nine possible ways in which your lower leg might nevertheless move upwards towards the plane of the table: (1) Someone strikes your knee in the appropriate place, producing the patellar reflex, and your leg jerks upwards; (2) Someone vastly stronger than you pulls your leg up, despite your valiant resistance efforts; (3) Someone threatens to kill you unless you raise your leg; (4) Someone surreptitiously slips a powerful hallucinogenic substance into the beverage you are sipping and you hallucinate that a rabid dog is about to attack you, an attack you try to ward off by kicking your leg at the hallucinated canine; (5) Someone has given you the post-hypnotic suggestion that you should raise your leg when the hypnotist snaps her fingers, and when she does so, you raise your leg; (6) You have an unconscious

---

\(^\text{12}\) I am borrowing here from John Maynard Keynes's famous quote from his General Theory of Employment: "[p]ractical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back." OXFORD DICTIONARY OF MODERN QUOTATIONS 120 (Tony Augarde ed., 1991).
hatred of authority and when a superior enters the room, you kick out as a sign of disrespect, believing falsely that this superior deserves it; (7) An impulse to raise your leg arises and, without thinking, you raise your leg; (8) You simply cannot understand other-regarding reasons for not doing what you want to do when you want to do it, and you now want to raise your leg, so you do; (9) Finally, someone offers you something you want even more than not to raise your leg if only you will raise it, so you do.

The first two cases are uncontroversially cases in which action is lacking: in case one an irresistible neurological mechanism is at work, and in case two an irresistible external force is literally physically compelling the movement of your body, much like an irresistible mechanism. Although something important was at stake for raising your leg because there was good reason not to, no one would either morally blame or punish you for the leg’s movement. After all, you did not raise it: there was no intention to raise the leg, no physically voluntary act. These are easy cases for just deserts and consequential theorists. You truly could not help yourself in either case and you would be excused. But these are not the cases for which a control excuse is required. The lack of an act defeats the prima facie case, whereas control excuses are meant to defeat established prima facie liability.

Arguments by analogy are powerful and there appears to be an almost irresistible impulse, based on what I term the “lure of mechanism,” to assimilate cases three through eight, and perhaps even the ninth, to the first two. In all of them, it is undeniably colloquial speech to claim, with varying degrees of success, that one cannot help oneself, that one’s act was “involuntary,” in all of them. The claimant thus hopes to borrow excusing force from the incontrovertible cases of mechanism and literal physical compulsion, but it is well to remember that all the other cases are not instances of physical mechanism or compulsion. In all, there is intentional, physically voluntary action.13 Perhaps an excuse should obtain in the others, but the meanings of “cannot help myself” and “involuntariness” are not literal, and the alluring metaphor of mechanism should not obscure the difference.14 If an excuse is to obtain in

14 But see John Gardner, The Activity Condition in Criminal Law, in RECHT UND MORAL 67 passim (Heike Jung et al. eds., 1991) (arguing that the analogy of
the latter seven cases, it must be one of excused action rather than the absence of action and excuses for action must be separately and not parasitically justified.

B. A Fundamental Psycholegal Error: Determinism/Universal Causation as a Control Excuse

As Michael Moore\textsuperscript{15} and I\textsuperscript{16} and countless others\textsuperscript{17} have recognized, the truth (or falsity) of determinism or universal causation does not entail that all human action should (or should not) be morally excused because all acts are nonculpably "out of control" in the sense, raised variously in the nine cases, of mechanism, physical compulsion, or, arguably, metaphorical compulsion. Most human movement is not literally compelled and most acts are not done under unusual constraints that might justify an excuse, such as to avoid threats of death or serious bodily harm or in response to hallucinations. Moreover, if determinism or universal causation were (or were not) true and the basis for an excuse, then everyone would (or would not be) excused for all action, an outcome inconsistent with the arguments for a discrete control excuse.

Some try to avoid the inference from these observations that determinism/universal causation is not the equivalent of compulsion by arguing for selective determinism.\textsuperscript{18} The thesis is that only some human action is determined and that only determined actions are out of control. But this argument is based on wildly implausible metaphysics. It is the theory of control that is doing the excusing work, not determinism, and there is no reason to suppose that determinism just happens to be on the job in just those cases we pick out for excuse on moral grounds.

An argument closely related to selective determinism and psychological to physical compulsion is virtually perfect, not metaphorical, and that the psychologically compelled agent does not choose).

\textsuperscript{15} See Michael S. Moore, Causation and the Excuses, 73 CAL. L. REV. 1091, 1128-48 (1985).
\textsuperscript{17} In recent decades, see, most notably, Peter Strawson, Freedom and Resentment, in FREE WILL 59 (Gary Watson ed., 1982).
equally implausible is based on an act's degree of causation. The more caused an act is, the more it allegedly is compelled. What reason is there to believe, however, that degrees of causation are a metaphysical feature of the universe, rather than a feature of our current level of knowledge, and lack of it, about causation? Presumably all phenomena of the universe are caused by the sufficient variables that produce them, whatever may be the case about our understanding of the causes. Just because we can explain why Jane does what she does, but not why Joan does what she does, does not mean that Joan's behavior is less caused than Jane's. Moreover, why is causation the equivalent of compulsion, even if it is true that the universe metaphysically admits of degrees of causation? Most action can be fully causally explained, at least by using folk psychological accounts alloyed with ordinary notions of proximate causation, but most action is not compelled in the sense that leads to the conclusion that the agent was "out of control." For example, each year when I politely ask my criminal law students with brown hair to raise their hands to aid a classroom demonstration, all the brunettes do so. The unanimous hand-raising is not a random event; it is fully caused by obvious variables, but there is not a scintilla of lack of control among the students. They all deserve my "thank you" for their admirably helpful, fully voluntary cooperation.

Many claim, of course, that if determinism/universal causation is true, then moral responsibility is unjustifiable make-believe that cannot support punishment. Properly understood, however, this is not the claim of control excuse advocates, because they do not wish to excuse everybody. If correct, the more general argument would obliterate individual moral responsibility, rather than identifying a particular condition, such as lack of control, that should excuse only those who meet its criteria. And even if true, the reason to obliterate moral analysis is not because we are all nonculpably out of control in the relevant sense. To excuse all on that basis is precisely to smuggle back into the moral analysis that which the general moral argument renders invalid.

20 As John Monahan pointed out in a personal communication, some students may perceive that they really have "no choice," no matter how politely the request is put, but as Parts II and IV demonstrate, such perceptions do not mean that the students are out of control in any morally relevant sense.
Now, I and most others think the general argument is also unpersuasive, but that dispute need not be rehearsed here. It is sufficient to recognize, first, that determinism/universal causation does not entail lack of control in the relevant sense, and, second, that acceptance of the more general argument would require a complete restructuring of our sense of ourselves as responsible agents, our moral practices more generally, and, not least, our system of criminal justice. Determinism or universal causation does not provide a discrete basis for justifying control excuses.

C. Lack of Intention?

It is apparent that allegedly "uncontrolled" action is intentional, even in the most extreme cases in which morality and law alike hold that an excuse is fully justified. Remember, to begin, that we are considering cases of action (cases three through nine), not bodily movements resulting from irresistible mechanism or literal physical compulsion (cases one and two). Consider cases of duress, in which the agent is threatened with death unless she does the wrong thing. The agent compelled to act by such threats clearly acts intentionally, chooses to do the alternative rather than face destruction. The agent's opportunity set is wrongfully and drastically limited in such conditions and we would surely excuse her, but not because she lacked intent. It is perfectly appropriate to say that she intended to save her life. For further support, the American Psychiatric Association generically defines "compulsive behavior"—a classic out-of-control case—as "intentional" and "purposeful."

22 I am assuming that under some conditions duress or coercion operates as an excuse. Some contend that even if the balance of evils is negative, the agent's conduct is justified if yielding to the threat is reasonable. See, e.g., R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS (forthcoming 1994) (manuscript at 185-87, on file with author); ALAN WERTHEIMER, COERCION 166 (1987) (noting that the success of a duress defense may only require that the harm threatened not be significantly less).

23 AMERICAN PSYCHIATRIC ASSOCIATION, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 247 (3d ed. rev., 1987) [hereinafter DSM-III-R]; see also AMERICAN PSYCHIATRIC ASSOCIATION, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 423 (4th ed., 1994) [hereinafter DSM-IV] (outlining the specific criteria for obsessive-compulsive disorder and defining compulsions as conduct "aimed at preventing or reducing distress or preventing some dreaded event or situation," which clearly implies that the conduct is purposive).

The American Psychiatric Association will publish DSM-IV in late May, 1994. I am very grateful to Ms. Nancy Vettorello, M.U.P., Administrative Co-ordinator of DSM-IV, who graciously provided me with advance copies of the DSM-IV materials cited in this Article. Ms. Vettorello informed me that few of the diagnostic categories
This analysis can be extended to all the hypothetical cases raised above. The hallucinating agent is motivated by false perceptions and beliefs, which limit her perceived opportunity set, but she acts intentionally when she kicks out to fend off the hallucinated animal. Even the purely impulsive person, who raises her leg for "no reason at all," does so intentionally, albeit "thoughtlessly." And so on. "Out-of-control" action is not necessarily unintentional action.

D. Will, Free Will, and Control

A standard claim about people who allegedly lack control is that they do so because they suffer from some defect of the will or volition, or because they lack free will. I shall address these claims in order, beginning with a brief excursion into the concept of the will.

Nonreductive theories of action uncontroversially posit that people act for reasons that are rationalized by desire/belief sets. Human action is based on practical reason. But it is notoriously true that practical syllogisms are not deductive. A person may have a desire/belief set that seemingly should ensue in a particular basic action, but the person may not act at all. When the person does act, how do desires, beliefs, and intentions lead to the bodily movements that we call voluntary acts? This is the mystery that the theory of the will or volition seeks to explain. In brief, an "operator" is necessary to get us from here—desires, beliefs, and intentions—to there—a bodily movement that will successfully (we hope) satisfy our desires.

Theories of the will or volition have waxed and waned in recent philosophy. Under the influence of Gilbert Ryle, for a short period the concept of the will was considered preposterous by the majority of action theorists, but in recent years, such a concept has become central to accounts of voluntary action. Some think that volitions are actions of the will; some treat the will or volition as simply another type of intention or trying. Michael Moore carried over from DSM-III-R "change the concept of the disorder, although there are some changes in the specific criteria." Telephone Interview with Nancy Vettorello (April 4, 1994).

Because DSM-III-R will still be the authoritative manual as of the planned publication date of this Article, I shall cite to both throughout.


25 See CARL GINET, ON ACTION 31 (1990) (noting that "will comes as close as any" verb to describing volition).

26 See ALFRED R. MELE, SPRINGS OF ACTION: UNDERSTANDING INTENTIONAL
argues that the will or volition is a functional mental state that translates desires, beliefs, and more general intentions into “basic” actions, including by resolving conflicts between intentions.\(^{27}\) This and similar functional accounts emphatically reject equating volitions with wants.\(^{28}\) In sum, modern theories treat the will in one fashion or another as an executory function.

Once one understands the meaning of the will or volition, it becomes apparent that the problem in control cases is not with the will, understood as an executory, functional state. The victim of a threat of death and the person motivated by beliefs generated by hallucinations or delusions are both able to execute the actions that will, respectively, save them from death or satisfy their crazily motivated desire/belief sets.\(^{29}\) Indeed, none of the nine cases above raised a problem with the will or volition, with the person’s ability to execute more general intentions. Even if an agent’s body is literally forced to move despite her strong desire to remain still, there is no defect or problem of the will. In the patellar reflex/mechanism case, the will was simply out of the picture altogether. There was no intention to execute. In the physical compulsion case, the agent’s will was perfectly capable of executing the intention to resist the stronger assailant. Agents can be physically forced or psychologically compelled to act against their desires, but the executory state is intact.

Now consider a case in contrast to the patellar reflex instance, in which the agent’s muscles are temporarily rendered helpless, unbeknownst to her. For example, suppose she is sitting with her

\(^{27}\) See Moore, supra note 5, at 113-65. Moore claims that the functional mental state that does the work is an “intention,” what he terms a type of “bare intention,” which “executes our more general plans into discrete bodily movements.” Id. at 121; see also Alan Donagan, Choice: The Essential Element in Human Action 21 (1987) (noting that an “acceptable theory of intention must show that human action can be explained neither in terms of beliefs and felt desires alone” but as “a general power to choose intellectual appetitive attitudes, whether they be choices to do now what you believe to be in your power, or wishes, or intentions”).

\(^{28}\) See Moore, supra note 5, at 120 (“[I]t is . . . not plausible to treat volitions as wants of any kind.”); Galen Strawson, Freedom and Belief 66-67 (1986) (noting that Immanuel Kant believed that “one possesses a will that is . . . a faculty distinct from desire” (citation omitted)).

legs crossed and is unaware that one leg has "gone to sleep." She now desires to get up and walk across the room. As she starts to rise, the leg fails to bear her weight and she falls, her desires thwarted, at least until the leg "wakes up." Once again, there was nothing wrong with the will. Her executory mental functioning worked perfectly, but an inadvertent, temporary circulatory problem caused her muscles to fail. Even in cases of so-called "weakness of the will," in which an agent acts contrary to her strongest desire, belief, or intention, the agent's action is clearly the intentional product of a well-functioning will.

In some of these cases, of course, we say colloquially that the agent's will was "overborne" in the sense that either the agent was forced to move or felt she "had to" act contrary to their preferences. But this is a misleading, metaphorical locution. As noted, volitions are not wants or desires: on the best theory, they are a species of intention. In the cases of no act and action alike, moving or acting contrary to other desires, beliefs, or intentions does not entail a problem with the will. Nonetheless, for various reasons some people undeniably seem to lack self-control, either more generally or in specific contexts. These people find it more difficult to behave themselves and are more disposed to offend than others who are better controlled. Still, the problem is not a defect in the will as an executory state of bare intention. The problem lies elsewhere.

One possible exception to the conclusion that out-of-control agents have intact wills might be cases in which there is a duty to act and the agent wants to do his duty, but he is psychologically "paralyzed." I have never encountered a judicial opinion addressing this issue, nor, I suspect, does it occur often in ordinary life. Still, such a case is surely possible. Imagine a parent with a pathological fear of open spaces, so-called "agoraphobia." He totally encloses and child-proofs the yard of the house so that his toddler can safely play unsupervised in the yard. Despite his admirable caution, the toddler one day suffers some obvious, untoward event, such as a seizure, that requires immediate attention. The parent wants to

---

30 I state the object of the conflict in the disjunctive because the nature of the conflict is contested and I do not wish to beg the question against any position.
31 See MOORE, supra note 5, at 140-41 (noting that agents "intentionally do acts that flout their strongest desires").
32 See supra text accompanying note 28.
33 See infra part III.
rush out to help the child—there is no conflict of desire, belief, or intention whatsoever—but he is "unable" to, experiencing "paralysis." We can even imagine that by brute force of will he goes to the door and starts to go out, but anxiety and its psychophysiological concomitants cause him to faint. Our hapless parent plausibly suffers from a volitional defect that interfered with his desire to do his duty. Truly he could not help himself. But if we were to exempt him from responsibility, note that the basis would be lack of action, not a control excuse. Note, too, that this case is recharacterizable as an irrationality problem if, as seems plausible, the parent's groundless fear of open spaces is deemed irrational.

Sometimes it is said that the "problem" with agents out of control is that they lack "free will." In almost all instances, however, this assertion cannot correctly mean either that there is a defect in the agent's executory mental functioning, or that action is out of control because it is determined or the product of universal causation. Often, I believe, the "unfree" will claim is used rhetorically to buttress an insufficiently supported conclusion that the out-of-control agent ought to be excused, because we all "know" that free will is a necessary component of, and perhaps sufficient for, moral and legal responsibility. This move creates a tautology, however, and a conclusory label, no matter how rhetorically powerful, does not provide justifications and criteria for excuse.

A more promising approach, although daunting, would be to enter the highly contested, technical free will literature to see what can be made of the claim that out-of-control agents lack free will. For example, one might say that only agents capable of rational self-reflection on their reasons for action possess free will and it is precisely this capacity that out-of-control agents lack. Or, one might say that agents lack free will if they act under certain constraints, such as externally imposed threats of serious harm or strong desires that the agent does not desire to possess—just the types of conditions that often lead to claims for control excuses. I will return

---

34 See, e.g., Randolph Clarke, Free Will and the Conditions of Moral Responsibility, 66 PHIL. STUD. 53, 54-55 (1992) (noting that an "agent who acts with free will directs her behavior on the basis of her view of the situation in which she acts"). But see Richard Double, How Rational Must Free Will Be?, 23 METAPHILOSOPHY 268, 277-78 (1992) (claiming that the quality of an agent's reasons has no bearing on freedom and denying the existence of free will).

35 See, e.g., HARRY G. FRANKFURT, Freedom of the Will and the Concept of a Person, in THE IMPORTANCE OF WHAT WE CARE ABOUT 11, 24 (1988) (noting that "[a] person's will is free only if he is free to have the will he wants"); see also JONATHAN GLOVER,
to such arguments in Part IV of this Article, but note that such arguments are, once again, not addressed to defects in the agent's narrowly conceived executory functioning or to problems that the truth of determinism might create. Rather, they are claims about the proper criteria for the moral responsibility of intentional agents—irrationality or coercion, for example; they are decidedly not about automatons or mechanisms.

In sum, trying to underpin control excuses in terms of will or volitional problems or lack of free will is likely to be inaccurate, confusing, rhetorical, or in its best incarnation, a placeholder for a fuller, more adequate theory of excusing conditions. The will and free will are not legal criteria, and agents in the criminal justice system would do well to dispense with employing them in responsibility analysis and attribution.

E. Irresistible Impulses?

Although an "irresistible impulse" is not a currently favored control excuse, it remains a criterion for legal insanity in some jurisdictions and it seems to exert a hold on the legal, mental health, and popular imagination. "Impulse control disorders" are an established category of mental disorders, some of which, such as "intermittent explosive disorder," kleptomania, pathological gambling, and pyromania, may produce behavior for which the agent will seek an excuse. Moreover, impulsive behavior is blamed for much criminal conduct and other antisocial behavior. Thus, there is reason to believe that attention to problematic impulses and

RESPONSIBILITY 61 (1970) (asserting that the three main types of excusable intentional acts are ones committed in response to threats, torture, or extreme need).

These arguments are addressed most specifically in Part IV.A. "Hierarchical" accounts of freedom, such as those deployed by Frankfurt, see supra note 35, are discussed infra notes 112-22.


impulsivity should shed light on control excuses. Once again, however, although the basic concepts appear clearly relevant to control difficulties, the potential for metaphor and confusion warrants caution.

Human beings incontrovertibly can be subject to momentary and apparently capricious passions that leave them feeling subjectively unfree and that seem to compromise their ability to control themselves. Such fleeting passions are often termed impulses and should be distinguished from cases in which such impulses are dispositional, which are usually termed impulsive or compulsive. Both impulses and compulsions are often thought to have the potential for coercive motivational force. Such observations, however characterized, are within the domain of common sense. The question is how these commonplaces bear on the justification for control excuses.

Note, first, that the impulses at issue are desires, fleeting and unconsidered desires to be sure, but simply desires nonetheless. If an agent acts to satisfy such a desire, doing so will surely be an intentional act executed by an undeniably effective will, and there is no reason to believe that universal causation or determinism plays a special role in such cases. The agent may have a strongly felt need to satisfy the impulse, but why is this different from standard cases of people desiring to fulfill momentary, strong desires? What would it mean to say that such a desire was literally irresistible? The lure of mechanism is clearly at work, but should be resisted. After all, why should a powerful desire—really, really wanting something—be assimilated to the patellar reflex? One possibility is that such impulses create a hard choice, but if so, hard choice analysis will do the work. A more likely possibility is that unthinking action in response to thoughtless or ephemerally thoughtful, momentary desires should be judged irrational in appropriate cases. But is such action better understood as irrational or as simply nonrational? In either case, rationality problems and not some supposed irresistible quality of the desire would be the ground for excuse when action is impulsive. Finally, it is famously the case that even if impulses do

---

43 When hard choices should excuse is addressed in detail in Part IV.A.
have coercive motivational force, it is impossible to differentiate "irresistible" impulses from those simply not resisted.

Impulsivity is a disposition or tendency to act with less forethought, or steeper time discounting, than most people of similar ability and knowledge. Despite the apparent consensus on this general definition, more specific criteria or descriptions have proved elusive. It is reasonable to assume, however, that at least some people who meet the general definition are dysfunctional, that is, suffer generally negative consequences, as a result of impulsivity. This assumption, too, is a commonplace and once again raises questions about why a disposition to act impulsively, as well as acting on an individual impulse, should excuse.

The dispositionally impulsive agent surely acts intentionally, with an effective will, and not under any particular influence of universal causation or determinism. Like the agent acting in response to an individual impulse, the dispositionally impulsive agent acting impulsively may experience a hard choice or act irrationally or nonrationally, but literal irresistibility will not be the operative variable to justify an excuse. Furthermore, an habitually impulsive person may have less ground for an excuse than an agent suddenly and unpredictably faced with an impulse: the dispositionally impulsive person knows that she is especially likely to act thoughtlessly and may therefore be held accountable for failure to take those steps that might remedy habitual impulsiveness or avoid those situations that facilitate it. One might argue in response, of course, that the dispositionally impulsive agent lacks the ability to plan ahead by taking such steps. If the agent genuinely did lack this ability, then blame would not be warranted. But the difficulty in planning ahead is likely to be a matter of degree, rather than an all-or-none quality. If so, the degree of difficulty required to excuse will be a normative judgment.

I believe that the general intuition supporting an argument for

---

4 See Scott Dickman, Impulsivity and Information Processing, in THE IMPULSIVE CLIENT, supra note 40, at 151; McCown & DeSimone, supra note 41, at 4.

45 See McCown & DeSimone, supra note 41, at 5; see also Dickman, supra note 44, at 153 (claiming that many of the inconsistencies in the impulsivity literature can be resolved by inferring the specific cognitive processes in which subjects differ).


47 The technology for assessing the irresistibility of impulsive behavior is addressed in Part V. I conclude that we lack the technology validly to measure the strength of impulses and impulsiveness.
excusing the dispositionally impulsive agent is not that desires are irresistible or that hard choice or irrationality exists. It is, instead, that the agent has the misfortune to possess a character trait that makes behaving oneself more difficult. Character rarely furnishes the basis for a legal excuse, however. The law assumes that people who are characterologically thoughtless, careless, pugnacious, excitable, cowardly, cruel, and the like can be expected to control themselves and should be held accountable if they violate the law. True, it may be harder for such people to behave, but the law assumes that they do not lack the ability to do so. Finally, if such characterological considerations were the basis for a control excuse, it would be because we decided as a normative matter that certain prophylactic personality traits were necessary for responsibility, not because the desires of characterologically disadvantaged agents were uniquely "irresistible."

F. Choice, Hard Choice, and Control

Some claim that responsibility resides in the ability to choose, an assertion that seems to support excusing out-of-control agents because, it is alleged, they have no choice. Philosophers of mind and action dispute the precise contours of choosing (understood as an agent’s mental act), but the technical intricacies of the concept are not central to the ordinary language notion that might justify control excuses. Nonetheless, the concept of choice so employed is ambiguous. Understood as a mental act, sometimes “choosing” seems to refer to the act of deciding between (at least two) alternative courses of action (or nonaction). Other times, choice as a mental act seems to be synonymous with acting intentionally (“I chose to go out for ice cream”). Finally, choice sometimes refers to a feature of the agent’s world that might be described as the

48 The array of such traits and qualities that makes behaving oneself harder and easier is discussed in Part III.


50 Compare MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 165 (1987) (noting that “intention is tied to further reasoning and action in ways in which choice need not be: that is why we do not intend everything that is an element in what we choose”) with MELE, supra note 26, at 140-41, 152 n.16 (noting that when deciding to perform an act, one also has the intention to perform the act). Such dispute about “choice” is unsurprising because the contours of all mental furniture are similarly contested.
alternative courses of action that were available ("I want ice cream, but none is available in my location or any other place I can go, and thus I have no choice about whether I shall have ice cream today"). Let us consider these ordinary uses of choice to understand why lack of choice or opportunity is an inaccurate or potentially confusing justification for control excuses.

Neither mental act usage is promising as a foundation for control excuses. As a general matter, out-of-control agents seem unproblematically able to choose between alternatives. If there is a gun at one’s head, one may find it exceedingly easy to choose to accede to the wrongful death threat. Similarly, the hallucinating agent has the mental capacity to choose to kick at the hallucinated dog rather than passively endure the feared attack. In the patellar reflex case, there simply was no choice as a mental act, but this is a case of no action and not an appropriate case for a control excuse. In the example of physical compulsion by the person who forces the agent’s leg up, the hapless agent has no choice, but not because the capacity to choose, as a mental capacity, was impaired. Indeed, the agent chooses between resistance and nonresistance, but, once again, the leg’s movement upward was not the agent’s act.

In some cases, a nonculpably ignorant or irrational agent may not be aware that a choice is possible. One might then claim that, at least in this instance, the agent does lack the ability to make a choice. Although this is not an implausible claim, note that it is entirely parasitic on other standard excusing conditions—ignorance and irrationality—which are doing all the work. In other cases, the agent might claim that the irresistibility of a desire deprived her of the capacity to make a choice. Again, such a characterization is plausible. But, assuming the validity of the claim about the strength of the desire, it seems more accurate to say, like the case of the agent acting under duress, that she was psychologically compelled to make a hard choice. She did, after all, choose to yield to the desire. Indeed, the strength of the desire made her choice easy, and if she struggled with conflict about yielding, this underscores the presence of the capacity to choose. The American Psychiatric Association’s generic definition of “compulsive behavior” as, inter alia, “purposeful,” “intentional,” and “designed to neutralize or to prevent discomfort or some dreaded event or situation,”51 again

---

51 DSM-III-R, supra note 23, at 245; see also DSM-IV, supra note 23, at 423 (worded slightly differently, but conceptually similar).
further supports the conclusion that the agent is able to exercise choice. Even if conflict remains "unresolved," agents are able to exercise and implement choice. In "irresistible desire" cases, then, some theory of psychological compulsion rather than lack of a capacity to choose is the possible justification for a control excuse. And if the terror of the choice set renders the agent "unable to think," such that no choice is possible, this is a rationality defect.

As a synonym for lack of intentional action, the other mental act notion, lack of choice as the basis for control excuses suffers from the same defects identified in the Sections above on intention and the will. Out-of-control agents choose their acts in this sense, so they do not lack choice in the same sense. In sum, lack of mental capacity to make a choice will not furnish general support for control excuses.

Lack of choice as lack of alternatives or opportunity is more promising, but this meaning can be both literal and metaphorical: to avoid the ever-present lure of mechanism, one must distinguish the two. On occasion, literally no relevant alternative action is open to an agent. The patellar reflex and physical compulsion examples are just such cases that defeat the prima facie responsibility requirements, which include a voluntary act. In contrast, some alternatives are available in all the out-of-control cases.

Those wishing to draw the analogy to examples of no literal choice claim that the agent had no "real" choice, or no reasonable choice. Indeed, we talk this way colloquially all the time. In brief, a hard choice is equated with no choice. For example, the person acting under sufficient duress has a choice—she might refuse to harm another, despite the awfulness of the threat—but she is a nonculpable victim of a wrongfully imposed hard choice and we cannot fairly expect her not to yield. For another example, the hallucinating agent had the option of doing nothing, but her mental disorder deprived her of the relevant information necessary rationally to understand her range of alternatives—there was no attacking dog, after all. Even judged from her internal point of view, passively enduring the attack of a vicious dog is (all things being equal) not a reasonable option that we can require of anyone.

52 ISAAC LEVI, HARD CHOICES: DECISION MAKING UNDER UNRESOLVED CONFLICT 34 (1986).
53 I shall discuss issues of ignorance, irrationality, and compulsion as justifications for control excuses in Part IV.
Hard choice cases in which we cannot expect the agent to behave differently undeniably exist, but note that what does the excusing work is not a defect in the agent. Instead, we are making a moral judgment about when options are so constrained that it is simply unfair to require the agent to behave otherwise. It is not that the agent literally was forced to do wrong and thus literally had no choice. Rather, as a moral matter, we might excuse because the choice the agent faced was too hard. Finally, even if hard choice situations explain why some out-of-control agents might be excused, many allegedly out-of-control agents, such as the impulsive person or the person without other-regarding understanding, might not be in a constrained choice situation at all. Hard choice does not mean that the agent lacks the capacity to exercise choice and it fails to furnish a general justification for excusing out-of-control agents.

I conclude that although colloquial talk about lack of choice is commonly used to characterize many out-of-control cases, it is often inaccurate and potentially misleading, as when the lure of mechanism leads to the conclusion that no difference exists between cases of no literal choice and cases of hard choice. Out-of-control agents should sometimes be excused, but not because they do not choose to do what they do. These cases are better analyzed directly in terms of ordinary justifications for excusing conditions, such as irrationality and coercion, as I try to do in Part IV. Before turning to this analysis, however, I shall first address those variables that help us remain in control.

III. FLYING STRAIGHT

The desire to do evil to our fellow humans, animals, and the world around us is ubiquitous. Few people have the good fortune to be free of such desires, and some people are burdened with intense desires to cause harm. How do enough people manage to fly straight in the face of these nasty desires and thus make human society possible? In this part of the Article, I propose a set of "self-protective" variables or devices that predispose people to stay on the straight and narrow and without which straight flight is difficult. Considering these variables allows us to see the commonsensical kernel of truth when an agent claims that she was out of control. It is harder to behave well, to "control oneself," to conform one's conduct to the requirements of the law, if one is not favored by possessing the "right stuff." But this accurate commonplace of moral psychology must be distinguished from the misleading
analyses to mechanism that are used to support claims for control excuses. If morality and the law grant an excuse if an agent lacks one or more of such variables, the proper justification is that a requirement of moral agency is lacking, not that the agent is out of control in the metaphorically mechanistic sense.

Rather than presuming to present an exhaustive account of moral psychology, I hope simply and briefly to suggest the major categories. I do not offer a developmental account, but instead include the relevant variables, however they may have come to exist. I focus on intrapersonal rather than social variables: although the latter exert strong causal force on the former, they usually operate through personal agency, and, ultimately, responsibility for criminal behavior is personal. For convenience, I shall discuss the variables in absolute terms, but it should be remembered that the strength of all may range along a continuum. I assume that most adults have relatively stable, “characteristic” types and amounts of the self-protective variables, but their strength can vary at different times in a person’s life. All the variables also have negative potential, but I shall focus only on the protective possibilities. Finally, I do not try to assess how many of these variables must obtain or in what strength and combination to insure straight flight. That task is beyond anyone’s capability.

The first protective variable is self-consciousness, the ability to monitor oneself. Effective human action is in general impossible without it, because it provides continuous feedback that permits us to adjust and readjust our conduct appropriately to surrounding circumstances. Self-consciousness is a standard feature of waking life, but it can be weakened or obliterated by a variety of causes, including trauma, intoxication, fatigue, dissociative states, and others. When this occurs, behaving properly is difficult because the agent is not capable of the normal degree of awareness. It is possible, of course, to go wrong in many ways despite normal self-monitoring. For example, an agent might be fully aware of what she is up to, but misperceive its import for other reasons. Sometimes it is good to “lose oneself,” but without general self-consciousness, the chance to understand what one is doing and to fly straight diminishes.

Social variables, such as general economic conditions, the hardening of targets, the availability of weapons and victims, and the luck of circumstances generally can have a profound effect on crime rates. Causation is not an excuse, however, and as noted in the text, responsibility in our criminal justice system is personal.
Fear of consequences is a powerful protective variable. The fear may be of external consequences, such as the infliction of physical or psychological pain, or of internal consequences, such as the experience of unpleasant psychological states like guilt or shame. It is good to have an environment that attaches (in the appropriate amounts and ways) negative consequences to wrongdoing, and it is good to have a well-functioning conscience or superego.

Rationality also protects. I do not have an exalted or complicated notion here, but only the ability, in Susan Wolf's words, "to be sensitive and responsive to relevant changes in one's situation and environment—that is, to be flexible." It is the ability to perceive accurately, to get the facts right, and to reason instrumentally, including weighing the facts appropriately and according to a minimally coherent preference-ordering. Put yet another way, it is the ability to act for good reasons and it is always a good reason not to act (or to act) if doing so (or not doing so) will be wrong. Having preferences and goals that meet apparent social standards for rationality will also protect. Whether desires and ends can be evaluated by rationality criteria is famously in dispute, but if what you want is generally considered irrational, you are likely to get into trouble.

Having moral desires or a disposition to behave morally also decreases the probability that one will offend. Often we say of such people, usually with admiration, that they have moral or good characters. Situational variables can act powerfully to enhance

\[55\] See SUSAN WOLF, FREEDOM WITHIN REASON 69 (1990).
\[57\] See DAVID HUME, A TREATISE OF HUMAN NATURE, 411 (L.A. Selby-Bigges ed., 1958) (1st ed. 1896) ("'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger."). Robert Nozick agrees that no one has yet provided a decisive answer. See ROBERT NOZICK, THE NATURE OF RATIONALITY 139-40 (1993) ("At present, we have no adequate theory of the substantive rationality of goals and desires, to put to rest Hume's statement."). Nozick goes on to argue that progress can be made by providing formal criteria that rational goals would have to meet. See id. at 141-51. He provides no content, however. See also DEREK PARFIT, REASONS AND PERSONS 120-26 (1984) (asserting that a reason's rationality depends upon its underlying desire, and that desires can be rationally based). Others try to provide content. See, e.g., CHARLES M. CULVER & BERNARD GERT, PHILOSOPHY IN MEDICINE: CONCEPTUAL AND ETHICAL ISSUES IN MEDICINE AND PSYCHIATRY 35 (1982) ("An irrational desire involves both wanting to suffer some evil and not having an adequate reason for doing so."); NICHOLAS RESCHER, RATIONALITY: A PHILOSOPHICAL INQUIRY INTO THE NATURE AND THE RATIONALE OF REASON 92-106 (1988) (arguing that "appropriateness" is the crucial evaluative criterion for the rationality of ends).

I return to this issue in Part IV.

or counteract any character, but, holding situations constant, straight flight is easier if the moral pilot is an ace. It is of course possible to fly straight even if one is inundated with wrongful desires, but it certainly will be harder.

Moderate temperament or emotions are useful, too. For Aristotle this was one of the marks of virtue. Anger, envy, and jealousy, for example, come immediately to mind as normal and often appropriate emotions that can wreak havoc on one's moral life by motivating awful behavior if uncontrolled. Moreover, becoming "too emotional" notoriously interferes with rationality, which is itself a classic self-protective variable. Indeed, it may make sense to conceive of inappropriately extreme emotions as irrational.

The capacity for empathy and identification with others is helpful. Criminal behavior harms and wrongs others, and agents understand the import of this more clearly if they can empathize with the pain potential victims will suffer. Such empathy strengthens one's reasons not to victimize. Psychopaths have great difficulty flying straight.

Having at one's command various self-control strategies that are either on "automatic pilot" or motivated in the right circumstances can be a genuine helpmeet for avoiding the crooked path. These methods can range in sophistication and subtlety, from techniques informed by research, to the brute resistance we term "sheer

59 See OWEN FLANAGAN, VARIETIES OF MORAL PERSONALITY: ETHICS AND PSYCHOLOGICAL REALISM 260-75 (1991); KUPPERMAN, supra note 58, at 59.
60 See Peter Arenella, Convicting the Morally Blameless: Reassessing the Relationship Between Legal and Moral Accountability, 39 UCLA L. REV. 1511, 1587 (1992) (claiming that the moral agent must be able to include moral considerations in practical reasoning).
62 See ALLAN GIBBARD, WISE CHOICES, APT FEELINGS: A THEORY OF NORMATIVE JUDGMENT 36-40 (1990) (also noting that the point is controversial).
65 See, e.g., AINSLIE, supra note 42, at 130-44 (providing picoeconomic explication of four precommitment strategies); MARTIN SELIGMAN, WHAT YOU CAN CHANGE AND WHAT YOU CAN'T: THE COMPLETE GUIDE TO SUCCESSFUL SELF-IMPROVEMENT 147-222
willpower." When nasty desires threaten to crank up or when one wants to engage in behavior, say drinking, that one knows will facilitate those desires, it is good to have successful methods available, whether or not crude, to suppress the desires or to preclude the facilitating conduct.

With some hesitation, I include good judgment or practical wisdom, partially distinguishable from theoretical reason, as the final self-protective variable. Common parlance makes the distinction, but I hesitate still because I am not sure exactly what the commonsense notion is, and it deconstructs readily under analysis. Still, we use the concept all the time, and it seems to do the work that "pure reason" does not do. In any case, we all feel mightily complimented if others say we have it, and we are certainly better able to behave ourselves if it is true.

Although all the self-protective variables or devices are helpful and it is a misfortune to lack any of them, it is not clear that responsibility is contingent upon possessing all of them in the minimally necessary amounts. Virtually all accounts of responsibility include self-awareness and rationality, but others, such as the capacities for empathy and guilt, are more controversial.

When addressing whether it is justifiable to blame and punish someone for wrongdoing, however, it will be useful to consider what self-protective variables the person possesses (or can be held accountable for failing to develop) and to remember that possession of the variables is mostly a function of heredity and environmental factors during childhood and adolescence that were not determined by the agent. It is a normative question whether or not we wish to include the lack of a self-protective variable among our reasons to

---


66 See Peter J. Steinberger, The Concept of Political Judgment 1-8 (1993) (questioning the distinction between intellectual skill and practical wisdom). But see Paul B. Baltes et al., Wisdom and Successful Aging, in NEBRASKA SYMPOSIUM ON MOTIVATION: PSYCHOLOGY AND AGING 123, 134-39 (Theo B. Sonderegger ed., 1992) (noting that empirical research on wisdom is in its infancy, but that the concept appears best understood as "expert knowledge in the fundamental pragmatics of life permitting exceptional insight and judgment involving complex and uncertain matters of the human condition").

67 See generally Arenella, supra note 60, at 1544-608.

68 Peter Arenella has suggested that the capacity for self-management of one's character is a necessary prerequisite of moral agency. See id. at 1609, 1615 n.120, 1618 n.128. He does not specify the criteria for this capacity, however, so it is unclear if it is a distinct self-protective variable or an amalgam of some of those already discussed, such as self-awareness, rationality, and self-control.
excuse, but answering that question will be easier if we have some empirical sense of how hard it was for a person to fly straight without it.

Finally, it is important to distinguish the various factors that comprise the agent's ability to fly straight, and thus to conform to the law, from the unanalyzed and loosely expressed conclusions that the agent suffered from a defect of volition, intention, or choice. For example, an agent who acts in response to irrational belief/desire sets has trouble conforming to the law because she is *irrational*, not because her will is unable to execute her irrational, more general intentions. Indeed, the case only comes to the law's attention because the agent's will *was* able to execute her irrational belief/desire sets and to do harm. Difficulties in flying straight almost never result from "defects" of volition, intention, or choice, as we shall see in more detail in the next Part of the Article.

IV. LOST CONTROL? CONTROL EXCUSES EXAMINED

In this Part, I address a wide variety of common circumstances in which offenders claim that they nonculpably could not control themselves and therefore should be excused. I conclude that notions of loss of control are almost always parasitic upon other justifications for excuse and that the notion of loss of control unduly threatens to mislead or confuse legislators, criminal justice system participants, and the public, even in those situations in which loss of control appears to be a plausible characterization of the defendant's behavior. I begin with the two classic cases of control excuses—duress and so-called irresistible impulse, or as I prefer to term it, "internal coercion"—both of which are based on some theory of coercion or compulsion. I then move to harder cases: psychopathy; the influence of dynamically unconscious motivation; cases at the border of no action and excused action, such as dissociative states, that should illuminate both sides of the border; and cases of deprivation or "rotten social background." Part IV concludes with discussion of the implications of the analysis for suggested new defenses based on a wide array of purported syndromes, such as battered victim syndrome and the like.69

---

69 I do not separately discuss clear cases of no action because, as has been repeatedly stressed, they deny the requirement of a voluntary act and do not properly raise control excuses.
A. Lack of Control as Hard Choice: Duress and Internal Coercion\textsuperscript{70}

This Section begins with a discussion of the general criteria for coercion and then applies them to cases of duress occasioned by external threats and to cases of "internal coercion" caused by internal threats. The latter are variously fashioned as "coercion," "involuntariness," "irresistible impulse," "compulsion," "volitional," or "control" problems. In the remainder of this Section, I shall use these phrases interchangeably because they all address the same conceptual domain of excuse and there is no uncontroversial reason to prefer any of them.\textsuperscript{71}

The almost universal acceptance of excusing claims of duress and coercion in criminal and civil law testifies to the implicit assumption that an excuse on these grounds should sometimes obtain. That is, the law often accepts that there are appropriate cases for excuse when the person claims, "I could not help myself." Nevertheless, understanding of coercion is far less established than understanding of defects of rationality. Although rationality is a normative concept about which there is no consensus among philosophers, psychologists, and others, there is a rough, common-sense consensus about the meaning of "everyday" rationality and its place in practical reason. In contrast, no consensus about the meaning of "coercion" exists among "experts" or laypeople. Consequently, my strategy is to begin with a choice among competing models of interpersonal coercion, about which much has been written. I shall first apply the chosen model to cases of duress. Coercion claims based on mental abnormality and other causes of internal threats are usually "one-party" cases, however. I shall therefore apply the chosen two-party model to such cases to determine if conceptual and practical progress is possible.

Theories of coercion may be classified into two types—empirical and moral. The former rely on analogies to cases of physical compulsion and suggest criteria for when a nonphysically compelled

\textsuperscript{70} This Section shamelessly cannibalizes and alters an earlier treatment of the same subject. See Stephen J. Morse, Diminished Capacity, in ACTION AND VALUE IN CRIMINAL LAW 239, 250-65 (Stephen Shute et al. eds., 1993).

\textsuperscript{71} But see Culver & Gert, supra note 57, at 116-17 (distinguishing external coercion from inner compulsion because the compulsive agent cannot will not to will the compelled conduct in the appropriate circumstances). I find Culver and Gert's distinction somewhat obscure and unpersuasive for the reasons this Subsection provides.
actor is deprived psychologically of the ability to behave otherwise. Psychological pressure or some negative affective calculus are the variables that do the work. (Remember that determinism is irrelevant to such criteria because if determinism were true and always produced coercion, then all action would be coerced.) The claim is that psychological compulsion makes it literally impossible for the person not to perform the allegedly compelled action.\(^7\)

Moral theorists abandon the quest for an empirical test because they believe that it is technically impracticable and normatively undesirable. They claim that the analogy to true physical compulsion is inapt and that we lack the understanding and expertise to "measure" the ability to do otherwise in the absence of physical compulsion or threats, such as a gun to one's head, that can be commonsensically assessed. Therefore, most empirical conclusions about coercion, especially those based on purely internal causes such as mental abnormality, are simply that, conclusions, unsupported by evidence. Even examples that intuitively support the empirical view, such as impulse disorders and substance abuse, derive their force from conceptual and factual premises that are far weaker and more controversial than their proponents assume.\(^7\)

Moreover, virtually all general empirical work addressed to the problems are unpersuasive or marginally relevant.

Perhaps most importantly, moral theorists argue that intuitions about psychological coercion that depend on the analogy with mechanistic, physical causation are often inconsistent with our considered moral evaluations and thus the former should not guide the latter unless the intuitions are well supported and the considered moral evaluation is unjustified.\(^4\) The facts upon which the analogy rests are highly problematic, however, and as we shall see, coercion excuses do not depend on the presence of "pressure," and coerced agents do act for reasons: they are not automatons; their movements are actions. Further, we believe upon moral reflection that the ability to resist should and does vary according to the

\(^7\) See Gardner, supra note 14, at 67, 73, 79. Mr. Gardner also claims that the analogy to physical compulsion is virtually perfect and not metaphorical and that the psychologically compelled agent does not choose. See also Culver & Gert, supra note 57, at 116.

\(^7\) See, e.g., Herbert Fingarette, Heavy Drinking: The Myth of Alcoholism as a Disease (1988); Alan Schwartz, Views of Addiction and the Duty to Warn, 75 Va. L. Rev. 509, 514 (1989) (arguing that "addictive behavior is chosen, not compelled").

\(^7\) See Wertheimer, supra note 22, at xi (endorsing the view that "coercion claims are moralized, ... they involve moral judgments at their core").
circumstances, and yielding to the desire should excuse only in some circumstances. For example, we do not excuse a drug dependent person who commits armed robbery to obtain money to buy drugs, no matter how much internal pressure the agent alleges to have felt. The analogy to a literally irresistible mechanism is simply too far-fetched properly to persuade us that such serious wrongdoing should be excused.

I have been convinced by the arguments of the empirical skeptics and moralizers, such as Fingarette, Nozick and Wertheimer, that a moralized approach is conceptually and practically preferable, despite its own difficulties. Moreover, moral models dominate discussion of the problem. Using a moral model, then, the working criteria for a coercion excuse are as follows:

First, the person is subjected to an unjustifiable threat, that is, unjustifiable circumstances that will make the person worse off compared to some baseline, if she does not perform the wrongful act.

76 See United States v. Moore, 486 F.2d 1139, 1149 (D.C. Cir. 1973) (en banc) (denying a lack of control excuse to a heroin-addicted defendant charged with possession of heroin).
77 See Herbert Fingarette, Victimization: A Legalist Analysis of Coercion, Deception, Undue Influence, and Excusable Prison Escape, 42 WASH. & LEE L. REV. 65 passim (1985). As Fingarette notes in speaking generally about coercion, “what looks like a comment about psychological causality becomes ... a statement about the act’s normative status under legal standards of wrongfulness and reasonableness.” Id. at 88-89; see also FINGARETTE & HASSE, supra note 29, at 51-65.
78 See Robert Nozick, Coercion, in PHILOSOPHY, SCIENCE, AND METHOD: ESSAYS IN HONOR OF ERNEST NAGEL 440 (Sidney Morgenbesser et al. eds., 1969) (attempting to “clarify the concept of coercion” and recognizing that there is an irreducibly normative component).
79 See WERTHEIMER, supra note 22, at xi (arguing that principles of coercion involve moral judgments).
80 More detail concerning the virtues and defects of the moralized approach is found throughout this Section and in Part V, concerning the factual assessment and adjudication of control excuses. On the problems—and virtues—with Wertheimer’s moral model, see, for example, Jerome E. Bickenbach, Critical Notice, 20 CAN. J. PHIL. 577, 581-82 (1990) (describing Wertheimer’s book as “fascinating, readable, and persuasive,” yet criticizing it for Wertheimer’s insistence on “draw[ing] the contrast between moralized and non-moralized interpretations of coercion too sharply”); Tony Honoré, A Theory of Coercion, 10 OXFORD J. LEGAL STUD. 94, 95 (1990) (arguing that “Wertheimer’s analysis fits ‘hard choice’ better than ‘defective capacity’ cases, but does not fit either exactly”).
Second, performing the wrongful act and suffering the threatened consequences for failing to perform it are both aversive choices, but doing the wrongful act is an excusable alternative because it is unfair to require the agent to refrain under the circumstances.

Third, the person is not responsible either for placing herself in or for failing to avoid the circumstances that produced the hard choice. In other words, coercion exists when the actor is not physically forced to act, but unjustifiable circumstances produce a hard choice and the agent cannot fairly be expected to avoid choosing to do wrong. Even though the person has a choice among actions—that is, no superior force is physically moving her body—there is no “real” or “acceptable” alternative, and the person cannot be expected to act otherwise.

Why adopt this particular, moral test of coercion? Various writers have attempted to grapple with the criteria for coercion, but although virtually all adopt a moral model, no particular test has achieved any consensus among philosophers. Each is subject to challenging counterexamples that expose fuzziness in the concepts, such as the difference between threats and offers, or that lead to results that seem undesirable. Consequently, one must simply make a choice, and the test chosen is, I believe, at least consistent with the various duress and coercion criteria in criminal and civil law.81 Moreover, the first criterion distinguishes threats, which are usually thought to decrease choice and increase coercion, from offers, which are usually thought to increase choice and decrease coercion, even if the offer is one the offeree “can’t refuse.”82 Some would argue that offers and mixed cases (“throffers”) can also be coercive, but this is a minority view that would tend to excuse people most observers wish to hold responsible. The threat must be unjustified because, within a moral model, a person should yield to a morally or legally sanctioned threat. Finally, even if unjustified threats are a necessary element, setting the baseline is itself problematic. As Wertheimer explains, the baseline can be statistical, phenomenological, or itself moralized.83 This last approach is best, because it

81 See, e.g., MODEL PENAL CODE § 2.09(1) (stating the elements of a duress defense).
83 See WERTHEIMER, supra note 22, at 206-11.
will yield a test for moral and legal excuse that does not collapse into conventionalism or unbridled subjectivity.

Turning to the second criterion, the threatened consequence and the wrongful act must both be aversive outcomes, or the person is not facing a hard choice. If the consequence is desired, the circumstance is no threat, and if the person desires independently to perform the wrongful act, then again there is no hard choice because the threat does not furnish the motive for the wrongful act. For example, the drug user or gambler who uses drugs or gambles primarily for pleasure and thrills and not to avoid anxiety or depression does not face a hard choice. But how hard must the choice be to warrant an excuse? This is the question that morality and the law must decide according to the circumstances of each context in which it arises. Once again, we are adopting a moral test that depends on social judgment, and the law generally requires very substantial threats before it will excuse agents for criminal conduct.

Finally, even if the agent is coerced according to the first two criteria, an excuse will not obtain if the agent was responsible either for placing herself in the situation or for failing to employ possible, resistant strategies. An agent who causes or fails to prevent the circumstances of her own excuse should not profit thereby.\textsuperscript{84} Although the third criterion is not specific to coercion,\textsuperscript{85} it is especially important when duress and mental abnormality furnish the coercive circumstances, because avoidance and resistance strategies are often genuinely possible.

\textsuperscript{84} See generally Paul H. Robinson, \textit{Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine}, 71 VA. L. REV. 1, 2-3 (1985) (proposing a new rule for treating actors who cause the conditions of their own defense and rejecting the "current law treatment" of such an actor with five types of common defenses).

\textsuperscript{85} The third criterion is not specific to a theory of coercion because the law often fails to grant an excuse or justification when the actor is responsible for causing the conditions that necessitated the excused or justified conduct. The law is blaming an agent for conduct that would otherwise be blameless because the agent was blameworthy earlier. For example, the Model Penal Code holds a drunken agent liable for recklessness even if intoxication deprived the agent of conscious awareness of the risk she was running. See MODEL PENAL CODE § 2.08(2). The justification is that the agent was allegedly reckless in becoming drunk.
1. Duress

Duress is seemingly the easiest case because an excuse is apparently unproblematic in at least some instances. Remember the generic criteria: the defendant is faced with a threat of suffering wrongful, immediate harm to herself unless she acts wrongly, and the threat is of such a nature and degree that a person of "reasonable firmness" would yield to it under the circumstances. Moreover, the defendant cannot be responsible for placing herself in the situation in which the threat would arise. The common law was more explicit and limiting: the threat had to be of death or serious bodily harm, and duress would never excuse homicide, no matter how dreadful the threat. The classic example, of course, involves a threat of death unless the defendant commits some other crime. Under such conditions, it would not be fair to expect the...

---


87 MODEL PENAL CODE § 2.09(1).

The common law's limitations threaten to obliterate duress as an excuse. First, because homicide is worse than all other crimes, death threats will always justify other crimes. Second, serious bodily harm either is arguably a greater evil than many serious crimes, once again producing a justification for committing the other serious crime, or it is not sufficiently bad to cause a person of "reasonable firmness" to yield to very serious crimes and no excuse will obtain. In sum, the general justification of necessity will obtain in those cases in which the balance of evils is positive, as it will always be in nonhomicide crimes occasioned by death threats or in cases of crimes less evil than suffering serious bodily harm. For example, in one well-known case, a gambler unable to pay his debts was threatened with serious harm to himself and his spouse unless he aided his gambling "creditor's" insurance fraud scheme by filing false insurance claims. See State v. Toscano, 378 A.2d 755 (N.J. 1977). Bracketing, for the moment, the issue of whether the gambler was responsible for placing himself in the situation, insurance fraud is surely less evil than grievous bodily harm or death and necessity would obtain. Duress as an excuse will otherwise seldom obtain because one can never kill, and even serious bodily harm is unlikely to support a duress excuse in cases involving other very serious crimes such as rape, kidnapping, arson, and the like. A person of reasonable firmness will be expected to suffer significant bodily harm rather than to engage in such crimes. The primary counterexample is the famous prison escape case, in which an inmate escapes to avoid serious bodily harm. Many courts have been willing, under usually limited conditions, to let the issue of duress go to the jury in such instances. See, e.g., People v. Lovercamp, 118 Cal. Rptr. 110, 116 (Ct. App. 1974) ("This [duress] defense is one with severe limitations and it must be established by competent evidence in a trial where the testimony of witnesses is subject to scrutiny by the trier of fact."). The law of duress is quite unforgiving, however, and such cases are rare. I shall assume for purposes of discussion, or else there will be nothing to discuss, that modern common law might permit duress as an excuse in cases of homicide when the balance of evils is not positive.
agent to behave otherwise or face criminal liability, blame, and punishment. We do say that the defendant was “forced to do it” or “couldn’t help it,” but what do we really mean?

Note a number of features of the criteria. Duress doctrine implicitly adopts a ranking of the moral seriousness of doing various evils and only allows the excuse under limited circumstances of “objective” reasonableness that are determined according to the implicit ranking. The criteria are infused with moral balancing and expectations. Second, the agent’s action designed to avoid the threatened evil is decidedly intentional. The threat furnishes excellent reason to form and to act on the intention. Third, the criteria do not require or imply that the defendant must experience untoward or unpleasant psychological states or that there is anything wrong with the agent. Most people threatened with dreadful immediate consequences will surely feel fear or “pressure” to avoid those consequences, and extreme fear might deprive some agents of their rationality, but such mental or affective states need not be present to claim duress successfully. In contrast, no matter how much fear or pressure a defendant subjectively experienced, duress will not obtain unless the situation was one in which a person of reasonable firmness would yield. If a person of reasonable firmness would not yield, but the defendant was driven crazy by honest but unreasonable fear, then an insanity defense is the appropriate excuse. If the amount of fear is reasonable, it will be difficult to claim that the agent was irrational if she acts to avoid the source of the fear. Furthermore, the threatened agent’s will and ability to exercise a choice are unimpaired. Choice is wrongfully and extremely constrained by the threat, but the defendant has and makes a choice that the will executes quite effectively to avoid the threat.

The claim that the duress defendant “can’t help it” is metaphorical and moral, rather than literal. We believe that it is unfair to ask the agent who is threatened with substantial evil to take the consequences rather than to do even greater evil herself. The agent experiences a wrongful “hard choice”: do evil or suffer the threatened consequences. She can help it, but under the circumstances it would be unreasonable and unfair to require her to do so upon the pain of conviction and its consequences. When it would be unfair is not a matter of the agent’s individual subjectivity, that is, how hard was it for her, in all her particularity, not to yield. Rather, it is a moral conclusion society and the law reach about what we may justly expect of each other as responsible beings. We
simply expect those who are weaker, more cowardly, fearful or the like to buck up or face the criminal consequences, even though we know such people find it harder to do so than those more fortunate-ly endowed.\(^9\)

If the justification of the duress excuse is the "reasonableness" of the agent's yielding under the circumstances, the claim sounds suspiciously like a justification rather than an excuse. How can reasonable behavior be wrongful? We are here, as Kent Greenawalt characterized the cases, at the "perplexing borders of justification and excuse."\(^9\) The act is objectively wrongful because the balance of evils is negative, but the agent subjectively behaves objectively reasonably. Michael Moore argues, contra Greenawalt, that the perplexing problem is solved by recognizing that justification is dependent on the objective moral features of the act, independent of the reasonableness of the agent's reasons for action.\(^9\) Although the problem remains perplexing, neither characterization affects the analysis of duress offered here. Whether a reasonable but wrong agent should be excused or justified, no defense obtains unless the defendant reasonably yielded (or believed, in the case of self-defense or defense of others). The defense is not grounded in a psychological, mechanism-like claim about pressure that is independent of preexisting moral analysis about the seriousness of evils and of the reasonableness of expectations about proper conduct.

How should the law respond to the agent who yields when a reasonable person would not, but who claims that she was "psychologically powerless" to resist the threat? I have already suggested that if fear drives someone crazy or otherwise deprives them of rationality (for example, they are so frightened that they literally "cannot think"), then an excuse based on irrationality, not duress,

---

\(^9\) In a personal communication, Peter Arenella suggests that the differences between law-abiding citizens and offenders may be far greater than is commonly assumed. That is, many offenders may substantially lack so many of the self-protective variables that it might be much harder for them to fly straight. Thus, moralistic, retributive sentiments and consequent punishment may often be unwarranted and unfair.

Such caution before inflicting blame and punishment is salutary because our criminal law too often imposes excessively harsh punishments, but we must also be careful not to deprive offenders unjustifiably of the respect they deserve as responsible moral agents.

\(^9\) Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1897 (1984); see also WALLACE, supra note 22; WERTHEIMER, supra note 22.

\(^9\) See MOORE, supra note 5, at 177-83.
is perhaps appropriate. Note that diminished rationality vitiates one's usual self-protection and thus makes it harder to fly straight, but the claim is not then mechanism or its moral analogues.

But how should we respond to the apparently rational agent who claims that her individual psychological characteristics prevented her from acting reasonably because the generally insufficient threat she faced was irresistible to her? If her claim is true, how can it be fair to blame and punish her? What is the nature of the claim? The first response is that the claim is not about "pressure" or psychological constraint, but about the rationality of her emotion, fear, under the circumstances. We commonly and plausibly talk about and judge what it is rational to feel as well as what it is rational to believe. Cases of doing evil under great fear occasioned by normatively lesser threats are good candidates for characterizing the agent as irrational on the basis of her motivating emotions. The second possibility is to treat the case as one of internal coercion, although the threat is external. That is, the objective evil of the threat is not doing the work, as it is in unproblematic duress cases. Rather, the external threat arouses fear of internal states that underpins the argument for excuse. I shall analyze this argument in the next section of Part IV, but for now we should recognize that this argument does not support standard duress cases.

In conclusion, duress should excuse, but not for the metaphorical reasons we commonly adduce.

2. Internal Coercion

The more difficult issue is how to understand claims about coercion resulting from mental abnormality—including claims about unusual cowardice or weakness that allegedly causes a victim to yield to an otherwise insufficient external threat. Can the objective, moral model of coercion be applied to one-party, intrapersonal, cases? The problem of intrapersonal involuntariness may be characterized generally as follows: you want to do something that you know you should not do, but you feel like you must do it anyhow because the pain of not doing it will be unbearable. Put another way, you experience intense and unpleasant emotion that can only be alleviated by wrongful action. Although some positive

---

92 See Radden, supra note 56, at 133.
pleasure or at least the "positive" experience of "release" will accompany the wrongdoing, the primary motivation is the avoidance of dysphoria. Now, if you do not know what you are doing or that it is wrong, this is once again a standard rationality problem and casts little light on inner coercion. Suppose, however, that you know rationally that you should not perform the wrongful act because you correctly believe that it is wrong, but you feel that you cannot help yourself because it will simply be too awful not to perform it. For example, you know that you cannot afford to lose any more money gambling but feel that you must place that next bet. Or, an agent feels overwhelmed by intense rage that can only be alleviated by violent action. Or, a person who abuses drugs but knows she should not, robs to obtain money for the drugs and shoots up again because she is terrified of the pain of withdrawal. Or, an agent desires to have sexual contact with a child, even though the agent knows that doing so would be exploitative and harmful. Or, you want to yield to someone who threatens you with mild injury even though you know that it is an objectively insufficient reason to do even greater evil. Most of the classes of mental abnormalities that involve alleged pathologies of the will—the impulse disorders, drug dependence, so-called paraphilias, and compulsions—fit this characterization, as does the case of unrea-

---

94 Once again, if the primary motivation is positive or pleasurable, then there is no threat and the second criterion of coercion is absent. See supra text accompanying note 83. For example, it is alleged that most pathological gamblers love to gamble and that gambling is "ego-syntonic," at least until the later stages of the disorder. Richard J. Rosenthal, Pathological Gambling, PSYCHIATRIC ANNALS, Feb. 1992, at 72, 73. If such were the case, internal coercion would not obtain.

95 See RADDEN, supra note 56, at 133.

96 See, e.g., Richard J. Rosenthal & Henry R. Lesieur, Self-Reported Withdrawal Symptoms and Pathological Gambling, 1 AM. J. ADDICTIONS 150, 152 (1992) (noting that many pathological gamblers report various withdrawal symptoms, such as insomnia or headaches, when attempting to slow down or cease gambling).

97 The American Psychiatric Association provides the following "essential features" of "impulse control disorders": (1) failure to resist an impulse, drive, or temptation to perform an act harmful to the self or others; (2) increasing sense of tension or arousal before committing the act; and (3) experiencing either pleasure, gratification, or release at the time of committing the act. See DSM-III-R, supra note 23, at 321; DSM-IV, supra note 23, at 609.

98 DSM-III-R generically defines these disorders as marked by recurrent sexual urges and fantasies directed at nonhuman objects, suffering or humiliation of oneself or one's partner, or children, or other nonconsenting persons. The disorder exists only if the person has acted on these urges and fantasies or is distressed by them. See DSM-III-R, supra note 23, at 279; DSM-IV, supra note 23, at 522-23.

99 Compulsions are defined as disorders marked by (1) repetitive, purposeful, and intentional behaviors performed in response to an obsession, or according to certain
Do the generic two-party criteria apply to the analogous one-party case? First, make the simplifying assumptions that an agent is not responsible for her desires and that an external threat can sometimes arouse dreadful dysphoria. Aristotle would demur to the first assumption, but many would agree, especially if the desire were considered pathological or if the agent experiences the desire as alien to herself, as in cases of compulsive hand-washing. And let us assume further that an agent makes every effort to avoid those situations that elicit the problematic desire. Despite these efforts, however, the desire arises or a dysphoric state is aroused and the agent will presently feel much worse unless she behaves wrongly. Although the threatened dysphoria from nonfulfillment of an unwanted, “abnormal” desire and from the continuation of unpleasant mental or emotional states cannot sensibly be described as a wrongful or morally unjustified “threat,” the despairing desirer surely does not “deserve” to be threatened. Consequently, it is reasonable to consider the threat “unjustified.” (In contrast, if positive pleasure were the primary motivation for wrongdoing, fulfillment would be an “offer.”) In sum, assume that the agent faces an unjustifiable internal threat, however caused, and is not responsible for having the pathological desire or other unpleasant state, for placing herself in environments likely to elicit it, or for otherwise failing to attempt reasonable resistance strategies.

Again, the critical question is whether performing the wrongful action to avoid the threatened or continued dysphoria is excusable. On the moral view, if the wrong thing desired is small beans, then it may be reasonable to do it, rather than to suffer substantial

---

rules or in a stereotyped fashion; (2) the behavior is designed to neutralize or to prevent discomfort or some dreaded event or situation, but the behavior is not realistically connected to what it is designed to neutralize or prevent; and (3) the person recognizes that the behavior is excessive or unreasonable. See DSM-III-R, supra note 23, at 247; DSM-IV, supra note 23, at 418.

100 See AINSLIE, supra note 42, at 205 (describing compulsions as having “coercive” motivational force).

101 See ARISTOTLE, supra note 61, bk. III, ch. 5 (implying that our desires are voluntary and so we alone are responsible for them).

102 See MELE, supra note 64, at 26-29, 50-61.
dysphoria. Suppose, for example, that the compulsive hand-washer's desire to wash builds to a crescendo just as her spouse is telling her something terribly important and she rudely and insensitively leaves to go wash. The spouse would not like it, of course, but if he has any charity in him, he would excuse her. Or, suppose that the pedophile unlawfully possesses child pornography as a means of gratifying his unwanted sexual urges. An excuse might not be unthinkable. Or, suppose that an enraged, cruelly jilted lover spews despicable epithets at the rejecting other. We might very well forgive the cruel words. But suppose, in contrast, that the hand-washer's crescendo of desire to wash her hands peaks just as her spouse chokes on some food and will die without immediate assistance. Or, suppose that the frustrated pedophile has intercourse with a child. Or, suppose that a drug-dependent person can obtain the money for the next fix only by committing armed robbery or burglary. In the latter cases—the choking spouse, the molesting pedophile, and the withdrawing drug-dependent felon—the moral test would hold that the person must bear the dysphoria rather than cause dreadful harmdoing.103

Although the moral analysis of two-party coercion cases appears profitably applicable to one-party cases of "internal" coercion,104 the analysis is complicated. In the remainder of this Section, I shall try to unpack internal coercion claims further, demonstrating that most cases are not pure internal coercion cases of hard choice, but instead should be analyzed as rationality problems. First, note again that one-party coercion cases are not instances of physical compulsion, in which an external or internal physically irresistible cause, such as a much stronger person or a neuromuscular reflex, moves a person's body although the person does not intend the movement and may even try valiantly not to perform it. In these cases a person literally has no choice and has not "acted." In contrast, the coerced agent has a desire/belief set that rationalizes the bodily

103 See Patricia Greenspan, Unfreedom and Responsibility, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS: NEW ESSAYS IN MORAL PSYCHOLOGY, supra note 10, at 63, 71-73. Note that the subjective view would simply inquire whether the person was psychologically capable of acting differently. The agent would be excused if she was able to persuade that the potential dysphoria allegedly made her feel helpless in the face of the desire, no matter how dreadful the wrongdoing and even if she had an acceptable alternative. This point is discussed further infra text accompanying note 130.
104 See Greenspan, supra note 37, at 196-99. But see Culver & Gert, supra note 57, at 116-17 (distinguishing between one-party and two-party cases).
movement when the agent washes to avoid dysphoria, strikes out in rage, sexually molests a child, reaches once again for the bottle, lays down yet another bet at the roulette wheel, or unreasonably yields to an objectively weak threat. Because many wish to excuse at least some people who yield to strong, allegedly pathological desires, they analogize goal-directed, intentional actions driven by such desires to truly involuntary movements. By this analogy they hope to strengthen the case for excuse. Remember, however, that the use of the words "coercion," "compulsion," "involuntary," and "irresistible" in these cases is moral and metaphorical—it does not have the literal, material definition that obtains in cases of physical compulsion. It is simply a loose characterization of those circumstances in which we excuse those who behave wrongfully in response to the unjustified threat of pathological desires or other dysphoric states.

Second, as Part II demonstrated, the internally coerced agent acts intentionally and exercises choice when she acts to avoid dysphoria. To hold that the agent acts unintentionally or does not choose is confusing, loose talk that begs the important questions. Choice can undoubtedly be substantially and wrongfully constrained, limiting the person's alternatives in the circumstances, but the decision to act or not to act is nevertheless an intentional choice, even under the most constraining circumstances. Thus, deciding which constraints should excuse will require a moral theory about excusing. Consider Martin Luther's claim: "Here I stand; I can do no other." Although there was "pressure" and no "real" alternative for Luther, he certainly chose.105 Or, consider the following case: suppose you decide to add a room with bath to your house and entertain bids on the same plans from two contractors, A and B, who are equally skilled, equally reputable, equally likeable, and equally efficient. A bids $60,000; B bids $45,000. Remember that all things are equal. Do you act intentionally and exercise a choice when you choose B, as any rational person would? Of course you do, although you would rightly claim, when you turned A down, that you really had no meaningful choice. Note in this case that the absence of meaningful choice would not allow you to claim coercion and avoid paying B. This situation involves an offer rather than a threat, of course, and thus fails to meet the moralized coercion criteria, but it does demonstrate, first, that choice is involved, even

---

105 I have used this example for years, but recently discovered while doing research for the present Article that Galen Strawson uses the same example to make the same point. See STRAWSON, supra note 28, at 139.
when there is "pressure" and no "real" alternative, and second, that the absence of meaningful choice does not per se excuse. The internal coercion problem is not lack of intention or choice—it is intentionally yielding to an unjustifiable choice in the absence of acceptable alternatives.

The third general consideration about one-party cases is that the agent’s conduct in response to so-called irresistible impulses, including impulses produced by intense emotions like rage, is in important ways, rational—the agent acts wrongfully “on purpose” for the perfectly rational reason that she wishes to avoid seemingly unbearable dysphoria. In the case of some impulse disorders and compulsions, such as kleptomania, the desire itself may seem irrational, but satisfying the need to avoid pain is surely not irrational. Moreover, for many people affected by the so-called paraphilias, some impulse disorders, and drug dependence, satisfying the desire produces positive pleasure as well as the avoidance of pain, and seeking pleasure is surely a rational reason to form an intention. If the motive for satisfying the desire is purely pleasure, then there is no threat and no compulsion, no matter how strong the desire is.

If a person’s ultimate goal, like stealing for no reason, is properly characterized as irrational, craving for it collapses into a rationality problem once again. And, irrationality is the basis for excusing if threatening circumstances arising from internal circumstances prevent the agent from thinking rationally. Many of the cases we mistakenly or loosely term “volitional” or involuntary fall under these descriptions. Indeed, some would claim that all cases are like this, even in the absence of obvious irrationality. Suppose, for example, that the person is not rendered cognitively irrational by threatening circumstances. Imagine that a person is petrified but rational: she has her wits about her, but feels that she must kill because she is morbidly afraid of bodily injury. Either the morbidity of the fear is itself irrational, or the intensity of it makes the agent unable in any meaningful sense to weigh the competing alternatives. In either case, irrationality is the touchstone.

---

106 See Parfit, supra note 57, at 120-21 (using claustrophobia as his example).
107 See supra note 94. Observe, parenthetically, that in cases where the actor satisfies the desire by wrongful conduct both to avoid dysphoria and to seek pleasure, the test of reasonableness for yielding to the desire is complicated on either the moral or the empirical view of internal coercion.
108 Note that if this case is treated (wrongfully) as a coercion case, the empirical model would excuse and the moral model would not.
Fourth, there is no defect in the will or volition, even if a person has intense, irrational desires that cause great dysphoria. As Part II also discussed, problems of coercion are distinct from what might be termed "pathologies" of the will, especially if the will is conceived as a functional, executory state. The functional mental state that produces action successfully satisfying an intense, irrational desire is as intact as the functional mental state that produces action satisfying an equally intense, rational desire. The "problem," if there is one, is either an excusable hard choice or irrationality, not volitional defect.\footnote{Fingarette and Hasse make this point very clearly. \textit{See FINGARETTE \& HASSE, supra note 29, at 55-65.}}

Nonetheless, there is persistent confusion in the excuse literature between so-called volitional or will defects on the one hand and internal coercion and irrationality problems on the other, as explanations for why allegedly internally coerced agents should be excused. Will defects rarely exist and most of the cases are anyway not internal coercion cases. For example, in a recent article, Dr. Richard Rogers cites the case of a woman suffering from major affective disorder who, in the depths of her hopeless dysphoria, attempts suicide and the homicide of her children to "end their suffering."\footnote{Richard Rogers, \textit{APA's Position on the Insanity Defense: Empiricism Versus Emotionalism,} 42 AM. PSYCH. 840, 844 (1987).} Although Rogers treats this case as one of defective volition, note that her will effectively executes her general intention to kill herself and the children. The real problem, of course, is that the depressed mother's assumption that the childrens' suffering is somehow indistinguishable from hers is a psychotic, gross misperception of reality, as is her belief that her situation is genuinely hopeless. Although there are clearly rational homicides and arguably rational suicides, this case presents neither, and terming it a volitional problem, especially the slaughter of the children, achieves no gain in comprehension.

But we can recharacterize the case, of course. Focusing solely on the suicide, we could treat the threat of unbearable, unending dysphoria as meeting the first compulsion criterion, and then treat the suicide as an acceptable alternative under the circumstances. But most of us do not believe that suicide is the only real alternative—we know that even the most severe depressions are self-limiting, that most respond to various treatment modalities, and that virtually all severe depressions compromise the sufferer's ability
to think rationally about her situation. Few would claim that this person is rational about herself, her disorder, and her future. If her situation really was hopeless, then the case may represent an entirely rational suicide that poses no volitional problems. Moreover, on this recharacterization, the woman’s will operated most effectively to end her dysphoria: the bare intention to commit the basic act that caused death itself executed the more general intention to kill herself.

Rogers also raises the case of a person with major mania, but once again the problem that ultimately causes legal trouble is the person’s beliefs and perceptions about herself and the world. The manic person does not knowingly do wrong because elevated mood somehow impels her to do so. Rather the mood disorder distorts her perception of reality and the consequent rationality of her practical reasoning. In any case, the will once again effectively translates the grandiose, irrational desire/belief set into action. Indeed, it is difficult to envision a case in which the defendant was suffering from a severe mental disorder with marked “coercive” features, but was substantially rational. Virtually all cases that would justify acquittal by reason of insanity or partial responsibility mitigation demonstrate that marked irrationality infected the practical reasoning that motivated the criminal conduct.¹¹¹ Crazy beliefs and perceptions are the touchstone. Nevertheless, the confusion of irrationality and volitional problems persists.

The most sophisticated attempt to rescue a volitional theory that analogizes physical compulsion to internal coercion without collapsing into irrationality employs hierarchical theories of motivation, such as those most famously deployed by Harry Frankfurt.¹¹² The central notion is that we are responsible for actions only if they are produced by desires that we identify with, ratify by evaluating them according to higher order desires. For example, some argue that agents lack the ability to act differently and do not choose to act unless they are identified with, assent to,

¹¹¹ Cases of impulse disorders and related diagnoses may be exceptions, but these are probably best characterized as cases of irrational desires. And, if there are situations of purely impulsive, thoughtless conduct in which the agent is incapable of any form of reflective awareness about her desires—cases that might be termed impetuous among the normal or explosive disorder among the abnormal—these are clearly cases of irrationality by any reasonable rationality criteria. Cf. DSM-III-R, supra note 23, at 321-22; DSM-IV, supra note 23, at 609-10 (“Intermittent Explosive Disorder”).

¹¹² See, e.g., FRANKFURT, supra note 35, at 11, 58, 159.
or ratify their desires, as the hand-washer, pedophile, and drug-dependent person presumably do not.\textsuperscript{113} Although hierarchical accounts are attractive,\textsuperscript{114} these accounts are problematic, and the concept of identification seems to do little work in the justification of excuse.

As Gary Watson has argued,\textsuperscript{115} higher order volitions, understood as Frankfurt and he use the term, are just desires themselves, and there is no reason to make them the touchstone of deliberation or any other criterion for responsibility.\textsuperscript{116} Moreover, what seems to give them authority is that they are evaluative, they mark what we consider worthwhile. But one can fail to identify with what one values and behave in ways one does not value from a more general standpoint. Watson argues that defining an evaluational system just in terms of what one does without regret abandons an explanation of self-determination that is based on identification by evaluation. Watson concludes that the notion of identification is "elusive" and that defining it as a type of "brute self-assertion seems totally unsatisfactory."\textsuperscript{117}

Another difficulty with hierarchical theories of responsibility is that ratification or identification does not do the work for which it is designed. An intensely greedy person, who accepts greediness as part of herself, may feel “powerless” in the face of temptation, even if we consider these desires normal (albeit undesirable).\textsuperscript{118} By contrast, a person with unwanted but weak pedophilic urges may have the ability to resist temptation, even if these urges are rightly called pathological. The identification criterion would condemn the former and excuse the latter, but this appears to be a perverse result that needs far more explanation. If we assume that the “identified”

\textsuperscript{113} See, e.g., Gardner, supra note 14, at 74-79.

\textsuperscript{114} But see Richard C. Boldt, The Construction of Responsibility in the Criminal Law, 140 U. PA. L. REV. 2245, 2256-62 (1992) (criticizing such accounts for their individualistic emphasis and failure convincingly to resolve the alleged incompatibility of determinism and responsibility); John M. Fischer, Responsibility and Control, in MORAL RESPONSIBILITY, supra note 37, at 174, 178-85.

\textsuperscript{115} See Gary Watson, Free Action and Free Will, 96 MIND 145, 149-51 (1987) [hereinafter Watson, Free Action]. Many scholars using hierarchical accounts rely on an earlier article by Watson that attempted to modify and thereby strengthen Frankfurt’s account. See Gary Watson, Free Agency, in FREE WILL, supra note 17, at 96 [hereinafter Watson, Free Agency]. These scholars often fail to notice that Watson is now much less sanguine about hierarchical accounts.

\textsuperscript{116} Recall that Michael Moore, Strawson, and others reject the account of volitions as desires. See supra text accompanying note 28.

\textsuperscript{117} Watson, Free Action, supra note 115, at 151.

\textsuperscript{118} For a discussion of the “moneyphile,” see infra text accompanying note 124.
agent has the ability to resist, but the “unidentified” agent does not, then the ability to resist, not identification, is doing the work and the empirical assumption about resistance ability needs further support.\textsuperscript{119} It is not immediately apparent why identification is coterminous with the ability to resist. Nor is it apparent that a person faced with even a dreadfully hard choice produced by her own wanted or unwanted desires is not choosing. For example, the American Psychiatric Association’s definition of a compulsive behavior defines it as purposeful and intentional—the agent is hardly an automaton.\textsuperscript{120} Hierarchical accounts do not convincingly disprove the claim that the analogy of psychological compulsion to physical compulsion is metaphorical. And, finally, suggestions that the agent has no choice beg the difficult empirical and moral questions concerning human abilities and what the law and morality can demand when choice is unjustifiably constrained.\textsuperscript{121}

Perhaps most controversially, Nozick argues about hierarchical accounts that the conflict between one’s desires at different levels violates a formal rule of the rationality of desires.\textsuperscript{122} One can make conflicting first and second order desires consistent by modifying either. To use Nozick’s example, if an agent whose first order desire is to take drugs has a second order desire not to have such a first order desire, the agent can achieve consistency by abandoning either the desire for drugs or the desire not to have the desire. Indeed, it might be entirely rational to abandon the second order desire if one believes that abandoning the desire for drugs would be far more difficult. I conclude, in sum, that hierarchical accounts do not provide independent reason to believe that agents suffering from untoward internal states should be excused because they cannot help themselves.

Fifth, an enduring mistake in analyzing one-party cases is the belief that abnormal cognitions are somehow more coercive or compelling than normal cognitions. An agent motivated by crazy beliefs is classically irrational, however, and there is no need to

\textsuperscript{119} Cf. MELE, supra note 64, at 73-74 (rejecting hierarchical accounts of motivation because they do not generate general resolution of the paradoxes of self-control).

\textsuperscript{120} See supra note 23.

\textsuperscript{121} Also, consider a person with unfortunate desires, who may have tried without avail to change and who has finally accepted her unpleasant fate because she has no alternative. Should this person become an enhanced candidate for moral appraisal because she has “ratified” her desires? After all, there is no positive evaluation; there is simply “brute acceptance” because life provides no alternative.

\textsuperscript{122} See NOZICK, supra note 57, at 141-42.
resort to internal coercion analysis. But in any case, it is a logical error to believe that mistaken perceptions and beliefs, whether normally or abnormally generated, are more compelling in practical reasoning than accurate perceptions and beliefs.\textsuperscript{123} The delusionally mistaken belief of a person suffering from paranoia that she is about to be attacked and must use self-defensive force is no more “compelling” than the accurate belief of a police officer that she must use deadly force in justifiable self-defense. Both have the same survival desire and there is no reason to doubt that both experience these desires with equal intensity. The unfortunate person with paranoia is of course irrational and in appropriate cases will be excused on that basis. The deluded agent might not have attacked but for the crazy belief, but the problem is the irrational belief and not lack of self-control. There is a defect in the agent that makes it harder for her to fly straight, but it is not lack of self-control mechanisms, unless these are defined generically to mean anything that makes it harder to fly straight. Furthermore, if an agent simply believed, without any apparent reason, that she had to attack an innocent victim or suffer some inchoate, dreadful dysphoria, the case would be clinically unlikely and one of irrationality.

Preliminary analysis of one-party cases suggests that most cases of pure internal coercion or compulsion that fit the hard choice model are better analyzed as irrationality cases than as hard choice cases. Moreover, they are not volitional problem cases. If an agent in a one-party case has trouble conforming, it is because irrationality interferes with the ability to fly straight and not because the will is overborne. Many would also treat cases of apparently pure hard choice—pedophilia, kleptomania, pathological gambling, or the coward who will commit any harm to avoid injury to self, no matter how slight—as rationality problems, even in the absence of cognitive irrationality, for these are cases of arguably “irrational” ends. Nevertheless, because the irrationality claim is controversial in such cases, let us consider the allegedly pure internal cases in more detail.\textsuperscript{124}

\textsuperscript{123} Cf. Jerome C. Wakefield, Disorder as Harmful Dysfunction: A Conceptual Critique of DSM-III-R’s Definition of Mental Disorder, 99 PSYCHOL. REV. 232, 237 (1992) (noting that reasonable beliefs and paranoid delusions can generate similar emotional reactions).

\textsuperscript{124} Practical concerns that may also be a reason not to adopt a hard choice analysis for excuse in internal coercion cases will be discussed in Part V.
How should morality and the law respond to a case of pure internal coercion—that is, a person who uses rational but wrongful means to avoid dysphoria threatened by arguably rational desires. Can these cases be treated as based on mental abnormality and on that ground as justifying a moral or legal excuse? Can these cases ultimately be distinguished from rationality problems?

Most laypeople and many clinicians would probably treat pure coercion cases as instances of clear-headed akrasia—that is, normal weakness of the will—and would hold the agent fully responsible because she does not seem sufficiently mentally abnormal. Suppose that a generally law-abiding person is nonetheless exceptionally greedy—a monephile, if you will. If this person is faced with a tempting situation in which the theft of a large sum of money is easily accomplished with little chance of detection, she may steal. How do we explain this case? One possibility is that she was so overcome by her desires that she failed to think straight about the moral and legal consequences of what she was doing. If so, the excuse, if any there be, is once again irrationality. If an actor "loses control," that is, does something that she would not otherwise do, as a result of a cognitive glitch, this is a rationality problem. The alternative possibility is that the agent recognizes the reality of the situation in all its moral relevance, but is somehow unable to refrain from acting wrongly because she fears mounting dysphoria or the like. This is the classic case of internal coercion explained by hard choice.

Do we excuse the monephile? The usual answer is negative: monephilia is considered a character trait rather than a disorder, and we believe that an agent is responsible for her character and able to maintain both cognitive rationality and self-control in the face of the strong desires her character produces, even when tempted directly. How is this case distinguishable, however, from pedophilia or gambling? Simply referring to the latter as mental disorders rather than as character traits begs the crucial question. Are our "normal desires" up to us more than our "abnormal/pathological desires"? We are all in large measure the product of biological endowments and environments over which we had no control and many of our central desires are firmly estab-

---

125 See GEORGE ELIOT, ADAM BEDE 166-67 (Harcourt et al. eds, 1962) (1859) ("Why, yes, a man can't very well steal a bank-note unless the bank-note lies within convenient reach; but he won't make us think him an honest man because he begins to howl at the bank-note for falling in his way.")
lished well before we reach the age of genuine, independent moral reflection on those desires. Moreover, what reason is there to believe that it is more difficult to learn to control abnormal desires than normal or immoral desires\textsuperscript{126} of equal strength? And if we try to distinguish the cases on the ground that pedophilia and compulsive gambling desires are pathological and irrational and monephilia is not—as is implied by the locution, "abnormal desires"—then we have redefined the problem once again as a rationality problem.

Another approach is to suggest that the desires of the pedophile or pathological gambler are necessarily stronger than the monephile's desires and that nonfulfillment will produce correspondingly greater dysphoria than in the case of the monephile. But this will not work either. There is simply no scientific or clinical evidence that "abnormal" desires are necessarily stronger than "normal" desires and thus that abnormal desires uniquely threaten unbearable dysphoria and produce a consequently harder choice. The monephile faced with an unattended pile at the ready may feel as much "pressure" as the pedophile unwittingly left alone with an attractive child. An extraordinarily strong desire for power, fame, or wealth motivates people to diverse, unseemly conduct, and for some people, pedophilic and other allegedly abnormal urges are mild and avoidable, even under the most devastatingly tempting circumstances. What is the relevance of the source of the desire except that some are "abnormal," that is, irrational, thus collapsing the analysis into the rationality domain once more? If desires or ends conceptually cannot be irrational per se, providing a principled way to distinguish these cases is difficult.

A final attempt to distinguish desires that diminish responsibility from those that do not employs the hierarchical view of motivation discussed above, which requires for responsibility that the agent identify with or ratify her desires according to higher order desires.\textsuperscript{127} Although hierarchical accounts initially seem to present promising responses to compulsive states,\textsuperscript{128} for the reasons given earlier—difficulties with the identification concept, with

\textsuperscript{126} Here I am assuming that one can distinguish between abnormal and immoral desires. To the extent that one believes it is impossible rationally to desire immoral ends, then the distinction collapses. See generally RONALD D. MILO, IMMORALITY (1984) (discussing the typology of immorality). This point is discussed further in the discussion of the psychopath's responsibility in Part IV.B.

\textsuperscript{127} See supra text accompanying notes 112-22.

\textsuperscript{128} See Watson, Free Action, supra note 115, at 148.
whether failure to identify is genuinely the basis for excuse, and with whether inconsistent desires are rational—this approach is unlikely to solve the dilemma.

Perhaps the soundest approach is simply to define as irrational any extreme desire that can threaten unbearable dysphoria, no matter how rational it might be in milder forms, and to limit internal coercion excuses to such cases. Indeed, most of us think there is something more than a little wacky about wanting anything "too much." Now, how much is "too much" will of course depend on the circumstances, including social conventions. An extreme desire to end hunger in one's society is less likely to be considered irrational than an equally extreme desire to possess the finest collection of matchbook covers in one's neighborhood. (And, how one tries to satisfy the desire will be judged as a matter of instrumental rationality.)

How should we respond to cases of wanting something "too much," cases in which the value of the good sought appears to bear no plausible relation to the strength of the desire for it? First, are desires that excessive appropriately characterized as rational? And when people are motivated to act wrongfully as a result of such extreme desires, do we believe that they are capable of rationally weighing the situation? I do not have answers to these questions, but my hunch is that most people would conclude that neither extreme desire nor practical reasoning that includes such desire is rational. And in many cases involving such extreme desires, it is also probable that the ability to reason well in the face of the relevant temptation will be compromised substantially.

Finally, on the moral view, how can threatened undesirable subjective states ever justify a rational actor's wrongdoing? The moral, objective test does not ask an empirical, phenomenological question that requires an answer about an unknown level of ability to refrain. The expectation of reasonableness is not a psychological variable, but a moral standard, and we assume that all agents can refrain from wrongful conduct, albeit some with greater difficulty than others. Thus, if it would be unfair to require a person to refrain from causing harm—as in the case of person who acts in response to the threatening gunslinger—the law will excuse her even if she is capable of refraining. Conversely, as a moral matter, we

\[\text{129 Cf. Culver & Gert, supra note 57, at 111 (providing the criteria for volitional ability).}\]
simply expect people to bear significant harms before they will be excused for harming others. Indeed, most American jurisdictions provide a duress defense only if the defendant had been threatened with death or grievous bodily harm, and most provide no duress defense to the crime of murder. The reasoning in two-party cases is that only the most seriously harmful threats can excuse, and in many jurisdictions no threat excuses taking a life. Understanding why one-party, internal coercion cases should be treated differently is obscure. Consequently, to justify excusing all but the most petty crimes, an agent would have to demonstrate that extraordinary fear of dysphoria drove her to unlawful conduct. There is good reason to believe, however, that even in the most stereotypically hard one-party choice, that of the drug addict, the choice is probably not so hard that one could not fly straight rather than commit serious crimes.\textsuperscript{130} And, in virtually all cases in which a sufficiently intense fear or other strong feelings support an internal hard choice excuse, these feelings would surely result from irrational beliefs or perceptions or would compromise rationality to a substantial degree.

Observe that if we adopted a subjective, empirical model for internal coercion, we would excuse anyone who persuaded us that she acted for fear of dysphoria, even if the dysphoria did not seem objectively intense and she committed a horrendous deed. Still, much as some people might terribly fear even slight physical harms, others might have similar difficulty bearing mildly unpleasant emotions. "Pressure" is "pressure," whether its source is objectively justifiable or not. So, the empirical model is hard put not to excuse the genuinely fearful physical or emotional coward. One may object, however, that if the source of the dysphoria was weak and the need to avoid it so terrible, then the problem must be characterological lack of self-control. But this is simply another way of saying the person is a coward. In either case, the agent experiences an inability to refrain. And is the empiricist willing to hold people responsible for their characters? How, at the age of self-reflection and maturity, can an intense coward justly be expected upon threat of punishment to develop the courage to fight and to conquer the cowardice? If this expectation is unreasonable, the internal coherence of the empirical model requires that this person must be excused. But excusing in such cases would be a morally perverse

\textsuperscript{130} See DOUGLAS N. HUSAK, DRUGS AND RIGHTS 108-17 (1992) (using a hard choice model to conclude that "addicts" are not powerless to stop using drugs).
result based on a behavior assessment technology that we lack. Finally, if an agent is willing to do something terrible to avoid objectively mild sources of dysphoria, one suspects once again that the agent's ability to weigh the alternatives rationally was impaired.

In the end, do pure coercion cases exist that require excuse? Although I am sympathetic to claims that the rationality of desires or ends is difficult to assess, I am finally convinced, by malignantly circular reasoning perhaps, that it must be irrational to want to produce unjustified harm so intensely that failure to satisfy that desire will create sufficient dysphoria to warrant an excuse. Moreover, in a very small class of cases, such as kleptomania or necrophilia, the agent's goal—described as theft for no reason or sexual desire for the dead—may simply seem unintelligible or "inappropriate" for any rationally motivated human being. The justification for the excuse in all these cases is then irrationality, not hard choice. Even if clinicians routinely consider what they perceive to be volitional problems in their clinical practice, it does not follow that the law must adopt a conceptually misguided excuse. As Joseph Livermore and Paul Meehl argued in their justly celebrated article on the virtues of M'Naghten, a morally justifiable insanity defense based on purely cognitive considerations is feasible. Even if "pure" internal coercion cases provide theoretically independent grounds for excusing, the profound conceptual difficulties already considered and consequent assessment and implementation problems, to be discussed in Part V, suggest great caution before adopting, analyzing, and adjudicating these as cases of hard choice rather than as cases of irrational action.

B. A Hard Case: Psychopathy

The traditionally-conceived psychopath is firmly in touch with the reality of the environment, including the schedule of rewards and punishments that the law and ordinary folk will impose for various sorts of behavior, and is able to engage in successful formal instrumental reasoning. The psychopath surely acts voluntarily and

---

131 See Rescher, supra note 57, at 92-106. But see generally Joyce C. Oates, "I Had No Other Thrill or Happiness," N.Y. REV. BOOKS, Mar. 24, 1994, at 52 (implying that for some serial killers, their deeds seemed the only means to achieve meaning or satisfaction in life, but that for others the conduct is simply inexplicable).

intentionally. Nevertheless, the psychopath lacks empathy and conscience, traits whose absence surely predispose an actor strongly to selfish, antisocial, and perhaps criminal conduct. The question is whether psychopaths should be excused for their bad deeds. There is no hint of internal coercion or defect in the will in these cases. The psychopath “cannot help himself” or “cannot conform” only in that he lacks important self-protective variables that would enable him to fly straight far more easily. So, if an excuse is to obtain, it must be that the capacities for guilt and empathy are independent requirements of responsibility or that the rationality requirement encompasses them. After all, can it be fair to blame and punish someone who lacks such crucial self-protective devices, especially because these variables are not only products of heredity and early environment, they are also notoriously hard to acquire as an adult if the agent lacks the right stuff to begin with.

I confess to great ambivalence about the proper moral and legal response to the psychopath’s wrongdoing. On the one hand, the

---

1994] CULPABILITY AND CONTROL 1635

Although the validity of the diagnostic category has been denied and there is much disagreement about the modal characteristics of psychopathy, I assume for purposes of discussion (and believe) that such people exist. The modern understanding of the concept begins with Hervey Cleckley, The Mask of Sanity (5th ed. rev. 1982), and the best recent, empirical work is by Robert Hare and associates. See, e.g., Robert D. Hare et al., The Revised Psychopathy Checklist: Reliability and Factor Structure 2 Psychol. Assessment 338 (1990) (asserting that the revised Psychopathy Checklist is a reliable and valid instrument for the assessment of criminal psychopathy in male populations); Stephen D. Hart et al., Psychopathy as a Risk Marker for Violence: Development and Validation of a Screening Version of the Revised Psychopathy Checklist, in Violence and Mental Disorder: Developments in Risk Assessment 81 (John Monahan & Henry J. Steadman eds., 1994) [hereinafter Violence and Mental Disorder] (same). See generally Hare, supra note 63. The characteristics of the “traditional” psychopath described in the text are not necessary criteria for the seemingly related disorder, “Antisocial Personality Disorder,” which is defined more behaviorally by the American Psychiatric Association in DSM-III-R, supra note 23, at 342-46; DSM-IV, supra note 23, at 649-50. The correlation between “Antisocial Personality Disorder” and instruments measuring the more traditional concept is strong but not perfect. See Hart et al., supra, at 92-93; see also Robert D. Hare et al., Psychopathy and the DSM-IV Criteria for Antisocial Personality Disorder, 100 J. Abnormal Psychol. 391 (1991) (describing an alternative approach to that given by DSM-III-R for “Antisocial Personality Disorder,” namely the revised Psychopathy Checklist). Nevertheless, a lack of conscience and empathy would surely predispose a person to the types of antisocial conduct that are criterial for “Antisocial Personality Disorder.” But see Michael Hakeem, The Assumption that Crime Is a Product of Individual Characteristics: A Prime Example from Psychiatry, in Theoretical Methods in Criminology 197, 207-10 (Robert F. Meier ed., 1985) (arguing that DSM-III’s category of “Antisocial Personality Disorder” is a tautological failure, but also failing to discuss the narrower notion “psychopathy” discussed in the text or Hare’s research discussed in this note).
psychopath knows what he is up to, what the rules are, and what will happen to him if he is caught for breaking them. From this vantage, the psychopath seems undoubtedly rational and many wish to hold him fully responsible if the only claim for excuse is psychopathy. On the other hand, he lacks attributes that give people perhaps the best reasons not to harm others and thus that operate as powerful moral, emotional, and intellectual inhibitors of harm doing. Viewed thusly, the psychopath seems "morally insane," unable successfully to reason practically about moral issues, to include moral concerns among his reasons for action.

Both characterizations are correct, of course, so the question is whether the law should adopt a standard of rationality that is "thin," requiring only selfish feelings, "bare" cognitive knowledge of the world and its rules, and the ability to reason instrumentally in a formal sense. Or, in the alternative, should the law adopt a standard that is "thick," requiring moral content in addition. If one adopts the former, no excuse is necessary; if the latter, some degree of excuse, depending on the agent's degree of psychopathy, is appropriate. I have not yet resolved this point satisfactorily for myself, so I must be content with simply raising the issue. Finally,

1636 UNIVERSITY OF PENNSYLVANIA LAW REVIEW [Vol. 142: 1587

134 See, e.g., Pillsbury, supra note 49, at 746-47 (claiming that psychopaths are rational and should be held responsible, unless they lack selfish feelings, which is highly improbable); see also MODEL PENAL CODE § 4.01(2), which provides, for purposes of excusing responsibility on the basis of mental disease or defect, that the terms mental disease or defect "do not include an abnormality manifested only by repeated criminal or otherwise anti-social conduct." This section is arguably meant to exclude psychopaths from those who can properly claim legal insanity. See MODEL PENAL CODE (Official Draft and Revised Comments 1985) § 4.01 cmt. 4. In contrast, some have argued that the psychopath should be entitled to raise a legal insanity claim as long as the diagnosis is not based only on repeated antisocial conduct. See, e.g., State v. Werlein, 401 N.W.2d 848, 851 (Wis. Ct. App. 1987) (deriving from the state criminal code a legislative intent "to exclude from the definition of mental disease or defect those disorders that are manifested solely by repeated criminal or otherwise antisocial conduct").

135 Arenella, supra note 60, at 1511 (examining whether the criminal law offers a persuasive account of a defendant's moral culpability); Susan Wolf, Sanity and the Metaphysics of Responsibility, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS: NEW ESSAYS IN MORAL PSYCHOLOGY, supra note 10, at 46. Wolf writes that these characters are less than fully sane. Since [they] lack the ability to know right from wrong, they are unable to revise their characters on the basis of right and wrong, and so their deep selves lack the resources and the reasons that might have served as a basis for self-correction. Id. at 58. In other words they are morally irrational in that they lack the ability accurately to reason morally. See also Jean Hampton, Mens Rea, SOC. PHILO. & POL'Y, Spring 1990, at 1, 14-15 (arguing that "knowledge of the [moral prescription's] authority is central to our finding [agents] at fault").
accepting an excuse for the psychopath does not put society on a slippery slope that will inevitably lead to the claim that none of us is responsible for what we do because none of us is responsible for who we are. Once again, the predicates for responsibility are the existence of certain attributes such as rationality and the lack of hard choice, not the existence of causation. We are all caused, but we are not all irrational, faced with hard choices, or perhaps, lacking any moral sense.

A final, consequential point about excusing psychopaths might tip the moral balance. Excused psychopaths would be preventively committed for extraordinarily long periods of time for two reasons: psychopathy is refractory to change, and if personality alterations that would warrant release did seem to occur, we would have special reason to be unsure whether a psychopath's seeming personality changes were genuine or feigned. An excuse might be the warrant for lifelong incarceration, even if the crime charged were one that entailed only moderate or short prison terms. Whether lengthy quasi-criminal commitment for psychopaths is desirable depends on one's view of the balance between the virtues of incarceration for public safety and the defects of deprivations of liberty. But this outcome would surely obtain if the criminal law excused psychopaths.

C. Dynamic Unconscious Motivation and Control

Adherents of psychodynamic psychological theories claim that much of human behavior is caused by psychological motives that are dynamically unconscious, that is, prevented from reaching awareness because recognition of them would provoke dreadful anxiety

---

136 The Supreme Court has held that the Due Process Clause of the Fourteenth Amendment permits involuntary commitment of an insanity acquittee for a period longer than the acquittee might have served in prison if convicted. See Jones v. United States, 463 U.S. 354, 369 (1983) (“There simply is no necessary correlation between severity of the offense and length of time necessary for recovery. The length of the acquittee's hypothetical criminal sentence therefore is irrelevant to the purposes of his commitment.”). The psychopath would surely have to be released if he were no longer dangerous, see Foucha v. Louisiana, 112 S. Ct. 1780, 1784, 1788-89 (1992) (noting that post-insanity acquittal commitment is justified only if the person is both disordered and dangerous, and citing Jones, 463 U.S. at 356, 368-69), but, as noted in the text, this would be uniquely hard to determine about a psychopath.

137 I chose to write on this topic because it is another on which Michael Moore has written influentially and taught me much. See Michael S. Moore, Responsibility and the Unconscious, 53 S. CAL. L. REV. 1563 (1980) (questioning whether the existence of unconscious mental states should alter moral or legal assessments of responsibility).
and other unpleasant feelings. An example of such motivation might be the bank robber who robs to pay gambling debts, but who executes his crimes in a manner that virtually insures that he will be caught. A psychodynamic formulation of the causes of his action might include the hypothesis that the robber unconsciously feels unworthy and guilty and desires, without being aware of it, of course, to be punished. As a result, and again without being aware of it, he commits his robberies in an unnecessarily incompetent manner, guaranteeing capture, conviction, and punishment. According to the dynamicist, such unconscious motivation is ubiquitous; there is, so to speak, a "shadow" system of practical reasoning of varying rationality that accompanies and influences our conscious motivation.

Assuming the validity of such hypotheses, what is their bearing on control excuses? First note that dynamically unconscious motivation does not negate intention or choice. The hapless bank robber may not have been aware of the "real" reason he robbed the bank, but he surely chose to rob it and did so intentionally. Moreover, he was fully conscious in the legal sense because he did not rob during a dissociative state. If psychodynamic motivation produces a lack of intention in some hard to fathom manner or, more plausibly but rarely, it causes a dissociative state, then the absence of mens rea or dissociation is doing the work.

\[138\] There is reason to have more than reasonable doubts about the validity of such assumptions. See Adolph Grünbaum, Validation in the Clinical Theory of Psychoanalysis: A Study in the Philosophy of Psychoanalysis 3 (1993) (challenging Freud's and post-Freudians' "clinical methods of validating causal inferences"); Stephen J. Morse, Failed Explanations and Criminal Responsibility: Experts and the Unconscious, 68 Va. L. Rev. 971, 983-1018 (1982) (arguing that psychodynamic theory does not provide scientifically validated causal accounts for behavior and that no means exist to construct reliable and valid formulations for behavior). For the more sanguine view, see Richard J. Bonnie & Christopher Slobogin, The Role of Mental Health Professionals in the Criminal Process: The Case for Informed Speculation, 66 Va. L. Rev. 427, 433-35 (1980) (arguing that speculation and imprecision regarding inquiries into mental aberrations are not persuasive reasons for courts to reject them). Treatment of the relation between dynamically unconscious motivation and responsibility in the depth it deserves goes far beyond the scope of this chapter, but a brief sketch of the answer is possible. For a far fuller treatment, see the sources cited above in this note. For a recent attempt to integrate dynamic and cognitive accounts of unconscious mental processes, see Mick Power & Chris R. Brewin, From Freud to Cognitive Science: A Contemporary Account of the Unconscious, 30 Brit. J. Clinical Psychol. 289, 302-07 (1991).

\[139\] For a discussion of the reasons to excuse dissociated agents, see infra part IV.D. In contrast, the presence of dynamic motivation does not per se negate mens rea, and, in virtually all cases, claims that particular unconscious motivation in fact
CULPABILITY AND CONTROL

 Should dynamically unconscious motivation have excusing force? Here are two theories that might support such an excuse. The first is that an agent who is unaware of the “real” reason for conduct is not rational; the second is that dynamic unconscious motivation somehow compels the agent to perform the conduct so motivated. Consider our sad bank robber again. He was quite consciously rational: he knew the relevant facts about the world and his conscious reason for robbing the bank—to obtain needed money—was certainly rational, if not laudable. Moreover, his will translated his desires into action. Is knowing the entire set of causes for one’s behavior necessary for responsibility? All of us almost always lack full awareness of the present variables causally influencing us, whether they are of the dynamic type or not. If the presence of dynamically unconscious causes or other unperceived causes negated responsibility, no one would ever be responsible because such causes are always operative. Only if dynamically unconscious motives were distinguishable for these purposes would this theory have plausibility. But because dynamic motivation is ubiquitous, we would then need a further theory and method for distinguishing unconscious motives that render the consciously rational agent actually irrational from unconscious motives that did not have this effect. For example, we might try to distinguish rational and irrational unconscious motivation. Even if this were possible, which is entirely a tooth fairy hypothesis, the excuse would be irrationality, not lack of control, except in the extended sense that irrationality makes flying straight harder.

Let us try one more irrationality approach to excusing the robber. He performed the robbery in a way calculated to fail to satisfy his conscious desire for money. How could he have messed up so badly unless he formed conscious irrational beliefs, say based on inaccurate perceptions, about the circumstances of the robbery? Such an account is plausible, and psychodynamic psychological

negated mens rea will be simply incredible. No story about unconscious motivation, no matter how clinically or scientifically credible, should or could convince us that an armed person who walks into a bank and demands money from a teller at gunpoint lacks the mens rea for bank robbery.

140 See RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 205-09 (1980) (describing subjects’ inability to generate accurate causal explanations of themselves and their behavior); Richard Nisbett & Timothy Wilson, Telling More Than We Can Know: Verbal Reports on Mental Processes, 84 PSYCHOL. REV. 291, 242-46 (1977) (asserting that individuals are sometimes unaware of stimuli that significantly influence physical responses).
evidence might be used to buttress the claim that the robber was “unable” to form true beliefs. As always assuming the validity of the entire account, we must face again the problem of distinguishing incompetence from an inability to perceive accurately, to form true beliefs, and to reason well—in brief, to behave rationally. If this were possible, the excuse would be classically conscious irrationality.

Let us turn to the alternative, compulsion theory. How might dynamically unconscious causes “compel” conduct? Why are dynamically unconscious causes any more compelling than the other, myriad causes of behavior of which we are unaware? Causes, even if they are unconscious, are not excuses. Of course, it is arguable that self-awareness about one’s “true” motives, in addition to self-awareness about what one is consciously doing, makes it easier to control conduct. But empirical research suggests that becoming aware of dynamically unconscious motives, that is, achieving “insight” into repressed psychological contents, does not help people to change,1 producing doubt about this argument. And, the argument is not about hard choice, but about another attribute, self-awareness, that may be self-protective. Assuming that awareness of one’s motives is self-protective, should lack of such awareness excuse in general or in the case of (some? which?) dynamic motives? If so, it would excuse because we believe that the self-protective variable is so important that lacking it makes it too difficult to fly straight. But there is no reason to believe this in general or in the case of dynamic motivation in particular. The “compulsion theory,” if supportable at all, reduces to a standard “hard to fly straight” theory, much akin to irrationality claims. I conclude that as long as the agent is consciously rational and not constrained by a perceived, blameless hard choice, justification for an excuse is lacking.

1 For example, the lack of difference in therapeutic outcome that different psychotherapies produce suggests that “insight” is not the mechanism of change in those psychotherapies that rely on insight. See Mary L. Smith et al., The Benefits of Psychotherapy 85-126 (1980) (evaluating the efficacy of different types of psychotherapies); see also Grünbaum, supra note 138, at 167-228 (discussing the lack of probative evidence for psychoanalytic theory and therapy). This point must be distinguished, however, from the possible value of introspection for understanding self-conceptions that are not dynamically unconscious. See, e.g., J. Gregory Hixon & William B. Swann, Jr., When Does Introspection Bear Fruit? Self-Reflection, Self-interest, and Interpersonal Choices, 64 J. Personality & Soc. Psychol. 35 (1993) (arguing that self-reflection may foster self-insight).
D. At Action's Border: Dissociative States

Dissociative states pose vexing problems for understanding human action and its relation to culpability. Although consciousness is always partial, these states involve a division, disturbance, or alteration in self-consciousness that ranges along a continuum from normal to severe. Normal examples might be "highway hypnosis," the phenomenon of driving quite competently for some distance without, apparently, either concurrent self-consciousness or later memory of having done so. Pathological examples include fugue states, sleepwalking, or episodes of depersonalization, in which a person becomes detached from the usual sense of self and may feel like an automaton. Intense emotions like rage or severe stress can trigger such states, which, again, can range along a continuum of severity. Automatism or unconsciousness is the standard criminal law doctrine that responds to such states when they are sufficiently severe to warrant exemption from responsibility.

Should these doctrines be understood as negations of a voluntary act or as affirmative defenses that excuse? On the one hand, the dissociated defendant has been able successfully to engage in conduct demonstrating accurate understanding of the environment and goal-directedness, suggesting that the bodily movements are intentional actions. These are certainly not

---


144 The defense of legal unconsciousness, which implies no voluntary act, should be distinguished from the psychodynamic use of "unconscious," discussed supra part IV.C., which has no necessary implications for excuse and rarely negates an allegation of a voluntary act, which is sometimes confused with the legal doctrine. See United States v. Pohlot, 827 F.2d 889, 906 (3rd Cir. 1987) (succumbing to the confusion, but reaching the right result nonetheless).

145 For a good discussion of Anglo-American law's ambivalence, see ROBERT F. SCHOPP, AUTOMATISM, INSANITY, AND THE PSYCHOLOGY OF CRIMINAL RESPONSIBILITY: A PHILOSOPHICAL INQUIRY 71-85 (1991). Because everyone agrees that severely dissociated agents should be exempted from responsibility on one of the two theories, allegedly "practical" lawyers may wonder why it makes a difference. Here are three reasons: it is theoretically important and interesting; the allocation of the burden of persuasion is affected; there may be substantial differences in the post-acquittal treatment of the agent.

146 Dissociative states are often followed by amnesia, but later amnesia does not necessarily entail that the agent lacked awareness or full intentionality during the
cases of reflex or physically compelled movement. Nor are they cases of random bodily movements: some mental state is directing the bodily movements quite effectively. Consider the case of Huey Newton. After becoming dissociated as a result of the trauma of being shot in the abdomen in a conflict with police officers, Newton was able to wrest a gun from one of his attackers, to shoot one, to run away, and finally, to go to the emergency room of a nearby hospital. On the other hand, in dissociative states, consciousness is not fully integrated because the normal ability self-consciously to observe oneself, to be aware of and monitor oneself, is missing or severely diminished. The self-protective variable of self-awareness seems crucial because it enables us to perceive our conduct and to behave more adaptively by correcting ourselves. In moral terms, the self-awareness operates as a censor or self-inhibitor: its absence makes it hard to fly straight by facilitating "unthinkingly" immoral behavior. For example, intoxication has the effect of "weakening control" because it inhibits self-observation and censorship. Dissociated agents that nonculpably and substantially lack self-protective self-consciousness should not be held responsible, but is the reason a theory of act negation or excuse? Is action lacking because volition is absent, or is the agent unable to fly straight because irrationality interferes?

Michael Moore suggests that dissociated bodily movements are not voluntary acts because, although "complex routines requiring perception and readjustment in order to reach certain goals" are performed, "the execution is not done by the conscious will, for the conscious will is 'elsewhere.'" For Moore, then, these are genuinely cases of volitional defect. I am undecided about this issue. Because Moore standardly presents the strongest possible case for any position he adopts, let us examine his claim in detail, beginning with a brief reconstruction of it that addresses his view of the relation of morality, responsibility, action, volition, personhood, and conduct. One may be fully aware of conduct and later amnestic, and dissociated conduct may or may not be followed by amnesia.

148 Psychoanalytic theory postulates that these two functions of consciousness are superego functions, but one need not adhere to Freud's theoretical structural model of the mind to recognize the existence and importance of these functions.
149 For the argument that voluntariness can be treated as part of the mens rea requirement, as well as of the act, see GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART 11-15 (2d ed. 1961).
150 Moore, supra note 5, at 257 (latter emphasis added).
CULPABILITY AND CONTROL

consciousness. At the risk of some oversimplification, I believe the Moorean view can be expressed by the following six, related propositions, each of which he defends: (1) Moral responsibility is an attribute of persons and not of subpersonal agencies and nonpersonal things;\(^{151}\) (2) It is the person’s choice to do wrong that makes the person responsible;\(^ {152}\) (3) Only human actions express the choices of persons;\(^ {153}\) (4) Human action requires functional states called volitions, which are nonactional, bare intentions that execute more general desires, beliefs, and intentions by causing the basic bodily movements that satisfy the more general mental states;\(^ {154}\) (5) Volitions are true mental states of “whole” persons and not subpersonal functional states;\(^ {155}\) and (6) Personhood and action exist only if consciousness is present.\(^ {156}\)

Moore is surely right that action, personhood, and responsibility all require consciousness, but what kind and how much? Moore defines “consciousness” as the “kind of awareness we have as an experience.”\(^ {157}\) In Law and Psychiatry, Moore offers a more expansive definition: one is conscious of something if the agent has the ability to identify her own mental states and feelings based on special, nonobservational knowledge, termed “privileged access.”\(^ {158}\) The consciousness required is not stream-of-consciousness concurrent awareness, however. The voluntarily acting agent need not be actually aware of the bare intention to perform the basic act that will satisfy the belief/desire set. For example, habitual and other “unthinking” acts are executed by volitions. But awareness of one’s volitions must be either easily accessible from the preconscious or recapturable in principle from the dynamic unconscious.\(^ {159}\) “Accessibility” appears to be what Moore means by “the ability” to know one’s own mental states in the special way consciousness requires.

As Moore recognizes, responsibility attribution would be unfair (and perhaps inefficient) if an agent’s lack of self-consciousness

\(^{151}\) See id. at 51, 151.
\(^{152}\) See id. at 51-52.
\(^{153}\) See id. at 51-52.
\(^{154}\) See id. at 135-55.
\(^{155}\) See id. at 132-33.
\(^{156}\) See id. at 151-55.
\(^{157}\) Id. at 151.
\(^{159}\) See Moore, supra note 5, at 51-52.
makes it unduly difficult not to violate moral norms; it is fair to
deeem harmdoing immoral only if harmdoers have the ability to fly
straight. But exempting the agent from responsibility could
result either from negation of the act requirement or from an
excuse. That dissociated agents cannot fairly be held responsible
does not entail that self-consciousness, as Moore defined it, is a
criterion for action. That needs to be defined independently, unless
action as a natural kind is partly defined by moral norms, which
seems implausible. I believe, however, that Moore does not
adequately defend why the ability to be aware of one’s volitions in
the way he posits is required for action or personhood, nor does he
indicate how much consciousness is necessary. It is possible that
unconscious agents are not responsible, but are nonetheless acting
persons. If a “thinner” conception of the consciousness required
for action is possible, then a normative choice between the two
based on nonaction grounds is what morality and the law require.
Or, in the alternative, the possibility of a thinner account suggests
that the definition of action is instinct with moral norms. To
demonstrate that a thinner conception is quite possible, I shall
reject Moore’s challenge to “come up with some alternative theory
about the essence of actions,” and shall instead use his own
topography.

Consider the sleepwalker, a classically dissociated, “unconscious”
agent. As Moore admits, the agent performs “complex routines
requiring perception and readjustment in order to reach certain
goals.” The sleepwalker is substantially aware of herself and of
her relationship to the environment. Accurate perception and
feedback loops to guide behavior are present. Most important, as
Moore recognizes, the movements of the unconscious agent that

160 See id. at 48-49.
161 Moore cites three clues to why human action is a natural kind: our first-person awareness of active control, our “actish” phenomenal feel; our sense that there is a
difference between the actions and mere movements of others; and, our belief that
the difference between actions and movements is a keystone to our sense of ourselves
and to our morality and law. See id. at 134-35. Now, unconscious agents may lack
“actish feel,” but maybe they (or some of them) do not. Many conscious agents also
lack it, and this feel is anyway only a clue to, not a criterion of, action. The second
clue is obscure when applied to unconscious agents, and the third states only that
action is distinct from movement and does not define the criteria for action. As I
shall argue, one can fully accept the importance of the distinction, but adopt a
“thinner” conception of action.
162 Id. at 255.
163 Id. at 257.
cause harms appear to execute more general intentions. After all, it is implausible that the harms done are random goals: they must express the agent’s individual desires and beliefs. Some sleepwalkers raid the refrigerator and some, like Ms. Cogdon, axe-bludgeon their children to death. To execute a general intention requires that the agent must be aware at some level of the intention that she is trying to execute. So, why isn’t the state that executes the sleepwalker’s more general intentions a volition? Why isn’t the quite substantial awareness of the sleepwalker enough?

Moore acknowledges that the sleepwalker’s functional executory state is “volition-like,” is a less-than-full “volition,” and does cause “just those movements to take place that would achieve the objects of one’s desire and general intentions . . . , which is why [unconscious movements] look so much like actions.” He denies “full” volition because, apparently, the executory state is not of the “whole” person: “[i]t may well be that subpersonal agencies within us are achieving quite complex functions in these . . kinds of cases.” This speculation is consistent with Moore’s stipulative definition of volition as being a state of the “whole” person, but what evidence is there for it? What naturalistic account of the type Moore favors would suggest that subpersonal proto-actions are another natural kind or that nature has endowed us with functional subpersonal bare intentions to execute more general but still subpersonal intentions?

Moore contends that to “verify” whether a volition is a true mental state of the “whole” person rather than a metaphor for a subpersonal mental-state-like routine, one examines the evidence

---

165 Id. at 257.
166 Id. at 258.
167 Id.
168 Id. at 257. I take up below Michael Moore’s other answer, which is that unconscious intentions do not perform the other essential volitional function—resolving conflict.
169 Consciousness is so poorly understood conceptually and scientifically that it seems unduly optimistic to base important moral and legal questions on speculation about allegedly naturalizable psychological structures and functions. See supra note 161 and infra note 186. Compare Daniel C. Dennett, Consciousness Explained (1991) with Flanagan, supra note 142 (agreeing that consciousness is a natural phenomenon worthy of investigation, but disagreeing about its nature). For a description and analysis of the checkered history of the study of consciousness, see Flanagan, supra note 142, at 1-20.
from phenomenology and behavior. The sleepwalker’s behavior strongly suggests that a true intention caused the goal-directed bodily movements. Now the sleepwalker cannot tell you about her phenomenology while she is dissociated, but neither can the admittedly volitional person performing habitual action on “automatic.” It does not seem at all metaphorical to suggest that the sleepwalker’s intentions are those of a whole person, even though she is not fully self-aware of them on the occasion. The unconscious agent’s behavior is simply too “actish,” too complex, too goal-directed, too dependent on awareness of self and environment, to claim that the agent’s movements lack the essential qualities of actions. Dissociated action differs from consciously integrated action in important ways, but it is more parsimonious to think that both are essentially actions.

Moore has two, complementary answers to arguments that dissociated movements are actions: (1) “[c]onsciousness seems essential as part of our self-boundaries, so that if we (our conscious selves) are [unconscious], then we don’t will anything”;\(^{170}\) (2) volitions cannot perform their “resolving” function unless they are “responsive to all (or at least a fair sample) of what one desires, believes, and intends,” and unconsciousness prevents such responsiveness.\(^{171}\) The first argument is attractively derived from common ways of speaking, but as Moore always hastens to remind us, ordinary speech can mislead. Perhaps it does so here. Let us do a Moorean thought experiment about what emotional reactions a properly moral agent would and should have if, like Ms. Cogdon, while sleepwalking, she ever so effectively axe-bludgeoned her daughter to death.\(^{172}\) She could say colloquially that she didn’t do it and Moore would support her theoretically: because she was unconscious, she did not will her daughter’s death. But should she feel guilty as well as enormously sad and regretful? She must understand, as do we, that murderous intentions powerful enough to be executed were part of her psyche. Isn’t some feeling of guilt appropriate? But why should she feel appropriately guilty if she didn’t do it? Doesn’t this suggest that she did do it in some

\(^{170}\) Moore, supra note 5, at 258 (first emphasis added).

\(^{171}\) Id. In accord with the second argument is SCHOPP, supra note 145, at 136-49 (expanding the argument using Dretske’s theory of action).

\(^{172}\) Moore famously uses thought experiments about worthy emotions to justify a retributive theory of punishment. See Moore, Moral Worth of Retribution, supra note 10, at 212-16.
CULPABILITY AND CONTROL

1994

important way? Consciousness does seem part of our self-boundaries and it may be harder to fly straight if one is not fully conscious of one's self, but this does not entail that the unconscious agent's movements are not actions. This claim, even if correct, is as consistent with excuse as with act negation.

The second claim is even more promising because an unpacked account of the causation of action drives it. Nevertheless, accepting the "resolving function" argument does not necessarily entail negating action and, as a practical matter, I fear that this argument may prove too much. First, I am not sure that there is always conflict to be resolved when action occurs. Cases surely exist in which only one unproblematic action option presents itself in an immediate way. Robert Schopp provides a helpful example of an agent who is cold, wants to be warm by putting on a nearby sweater, and does so. Although other opportunities for action were available and there were opportunity costs involved in donning the sweater, it is trivially true to the vanishing point that "conflict" is "resolved" in this case. Thus, it is not certain that resolving conflict is an essential function involved in the causation of all action.

On the other hand, virtually all desires to cause harm will arouse conflict in nonpsychopathic agents because knowledge that harm will be done is virtually always a good reason not to implement one's desires. According to Moore, volitions cannot resolve unless they are responsive to a "fair sample" of what other beliefs, desires, and intentions the agent possesses. Robert Schopp argues some-what differently:

the actor who selects an action-plan in a state of impaired consciousness acts without the benefit of the causal force that would ordinarily be exerted by certain wants and beliefs that constitute reasons for acting in a certain manner. . . . [W]ants, beliefs, and decisions would produce actions, but they would not do so in the manner of ordinary human activity because the full array of background wants and beliefs would not be available to

179 In the course of making the second argument, Moore says that, "volitions are part of the hierarchy of bare intentions that resolve, all things considered, what to do now." MOORE, supra note 5, at 258 (referring to the discussion in chapter 6, at 137-49). This is a bit perplexing, however, because his earlier discussion of the inevitability of conflict and volition's resolving function does not include discussion of hierarchies of intentions or of how hierarchies would work. It is devoted to proving, quite successfully, that the resolving function is an intention, rather than a desire or belief. I await the next book for elucidation.

174 See SCHOPP, supra note 145, at 115.
175 MOORE, supra note 5, at 258.
the reasoner during the process of deliberation. . . . Deprivation of access to these wants and beliefs distorts the causal process by which the actor's mental states cause his selection of an action-plan, and therefore his behavior is not produced by his effort and determination in the manner of ordinary action.176

Both agree that impaired consciousness blocks access to the psychic materials ordinarily available to help the agent deliberate about what to do when conflict is present. The agent nonculpably does not have the usual self-protective access to the reasons not to cause harm. To use Moore's phrase, they have been "sealed off,"177 and the agent will consequently have trouble flying straight. Exemption from responsibility is surely plausible, but once again the justification is as consistent with excuse as with act negation. Moore claims that access is necessary to permit volitions to perform their resolving function and that action is absent without the operation of this function.178 Robert Schopp, in contrast, seems to suggest that action occurs, but that it is abnormally caused.179

First, note how normative both accounts are. How much is a fair sample? How much of the background wants and beliefs must be available? Determining the right amount will partly and perhaps entirely depend on moral norms. More importantly, the dissociated agent is not "paralyzed" by conflict. One possibility in situations of conflict is that the agent does nothing, is still, does no act at all because the concluding all-out propositional attitude simply cannot resolve the conflict. In contrast, the dissociated agent performs complex routines that appear to execute more general intentions. Moreover, there are always choices about conflicting means and the resolving function seems quite up to this task. For example, Ms. Cogdon might have performed the complex action of killing her daughter by any number of means, each of which had opportunity costs, yet all the choices were resolved by a concluding choice to axe-bludgeon. In a more complex case, say the Huey Newton example, it appears even more implausible to believe that the resolving function was not doing its work as Newton engaged in the extended course of conduct that killed the police officer and that concluded with Newton's arrival in a hospital emergency room. Finally, the countervailing considerations are not obliterated, but

176 SCHOPP, supra note 145, at 148-49 (emphasis added).
177 MOORE, supra note 5, at 258.
178 See id.
179 See SCHOPP, supra note 145, at 149, 152.
simply out of concurrent awareness. It does not seem implausible to characterize dissociation cases as ones in which conflict was resolved, albeit on “thin” grounds. Why isn’t this enough functional conflict resolution for action?

I believe that to decide whether dissociation negates action or excuses, psychopathy may provide a better analogy than the patellar reflex. The dissociated agent presumably has the usual countervailing reasons rumbling around somewhere in the psyche, whereas the psychopath, who uncontroversially acts, never has these reasons available at any level. Psychopaths are thus essentially “thin” moral agents. Moreover, if psychopaths have diminished ability to learn from previous punishment, they also lack the capacity to use ordinary prudential considerations to block harmdoing. If psychopathy warrants exemption from responsibility, it is primarily because the psychopath is “morally insane,” not capable of moral rationality, and to a lesser extent, perhaps, incapable of prudential rationality. The dissociated agent has moral reasons blocked off and, perhaps, prudential reasons too, although the Newton case gives one pause about the latter. In sum, if we justifiably believe that the morally “thin” psychopath acts, rather than claim that it is simply Ms. Cogdon’s body and not Ms. Cogdon who killed her daughter, it seems more plausible to say that it was a “thin” version of Ms. Cogdon’s ordinary self who killed.

I also fear that the resolving function argument proves too much because it would obliterate action in a substantial array of cases in which our best considered judgment is that action surely occurred. Many crimes against the person are crimes of “passion,” committed in heightened emotional states, such as fear and rage, that may seal off access to the ordinary desires, beliefs, and intentions that permit volitions to resolve the inevitable conflict by being properly responsive to those background factors.Sometimes agents may be characteristically prone to such untoward emotional states and yet do nothing to try to master them or to avoid situations in which they are aroused. In many other cases, however, environmental stimuli may “sneak up” on the well-motivated agent, or the agent’s ability to master untoward emotions on the occasion may be reduced by fatigue and other variables. An enraged, jealous agent who discovers spousal infidelity and kills immediately, in the heat of passion, may be as unable to have access to a fair sample of wants, beliefs, and desires as Ms. Cogdon was. Or, a grieving person who kills immediately in response to lesser provocation, but
in a state of extreme emotional disturbance, may also be disabled from normal action causation.

On Moore's account, action is plausibly lacking in these cases. Yet, at most, the law (and ordinary morality) responds to these cases with a "partial excuse," such as the provocation/passion doctrine that reduces intentional killings from murder to voluntary manslaughter or the Model Penal Code's even more forgiving "extreme emotional disturbance" doctrine.\textsuperscript{180} The law seems right in general, but I would argue that it does not go far enough. Action is present, but in some cases there is sufficiently extreme emotional disturbance to warrant outright acquittal, as in standard cases of extreme dissociation.\textsuperscript{181} In less extreme cases, including less extreme dissociation, the most appropriate response is a partial excuse. For example, sleepwalking cases may require outright acquittal or partial excuse, depending on facts we might learn about sleepwalking in general and about a particular sleepwalking defendant. There is no reason to believe that all sleepwalkers are dissociated to the same degree. For example, if we discovered that Ms. Cogdon was a "light" sleepwalker, might it not be appropriate to suggest that she killed in a state of extreme emotional disturbance and therefore should be convicted of a lesser homicide crime?

To conclude discussion of Moore's account of unconsciousness, consider his definition of consciousness: a disposition or ability to have access to the preconscious or to recapture dynamically unconscious contents.\textsuperscript{182} Do dissociated agents have this ability? Moore assumes, it appears, that they do not, because he believes that unconscious agents do not act. The truth is uncertain, however. Most dissociation cases, such as sleepwalking and fugue states generally, surely involve dynamically unconscious states, whether caused by the panoply of psychodynamic explanations or others. According to Moore, mental states must be recapturable only in principle, however difficult that may be, to establish that they are the person's.\textsuperscript{183} And, all that needs to be recaptured is the agent's general intention to kill, assault, or the like, because, as Moore sensibly notes, "conscious awareness... has better things to do

\textsuperscript{180} Model Penal Code § 210.3(1)(b).
\textsuperscript{182} See Moore, supra note 5, at 151-52.
\textsuperscript{183} See id. at 152.
than be squandered on the details of motor movement." There is no reason to believe that unconscious agents might not recapture their general intentions if exposed to various forms of psychological methods, perhaps including soul-searching. Ms. Cogdon and Huey Newton surely could in principle recapture their homicidal intentions. Even if dissociation is dynamically produced in the same way in different agents, remember that similarly dissociated agents will not behave the same. Sleepwalking refrigerator raiders and sleepwalking axe-bludgeoners have very different, preexisting general intentions that were available to be executed and that are in principle recapturable. On Moore's own account, unconscious agents may act.

Despite all the foregoing cautions about the claim that unconscious agents are blameless because they do not act, I am still undecided. It is a hard, close case that I do not think can be decided by a natural kind theory of action, which is in principle naturalizable and divorced from normative, moral considerations. Because the moral case is so close and uncertain, and despite the risk of undermining the moral importance of the act requirement, I am inclined to shift to consequential morality and for practical reasons to treat these cases as raising claims of affirmative defense. First, the outcome in either case is exemption from responsibility and outright acquittal, unless the crime charged includes a lesser, negligence crime. But negligence convictions are not appropriate because it is unreasonable to treat the severely dissociated person as capable of behaving as a reasonable person. The defendant who acted in a dissociated state is more like a legally insane actor than like an actor who harms as a result of a reflex movement or than like a rational defendant who made a mistake of fact that negated mens rea. In any case then, substantial injustice will not occur if dissociated defendants are acquitted by virtue of an affirmative defense rather than by act negation. Second, claims of dissociation are difficult to establish, easy to fake, and may be

184 Id. at 153.
186 Cf. H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 152-54 (1968) (arguing that it is just to punish negligence only if the defendant was capable of behaving reasonably).
187 Consider the dispute about whether the most exotic form of dissociative disorder, multiple personality, exists at all. See Harold Merskey, The Manufacture of Personalities: The Production of Multiple Personality Disorder, 160 BRIT. J. PSYCHIATRY
easy to raise in common cases of dangerous, violent actors whose conduct is touched off by rage and other strong emotions. Consequently, there are many reasons to place the burden of production and perhaps also the burden of persuasion on the defendant: justifiable acquittal will be rare, the defendant has the best access to the necessary evidence, and public safety may be unduly compromised by wrongful acquittals.

E. Deprivation or “Rotten Social Background”

Many claim that people who have suffered terrible lives or who have been brought up in seemingly criminogenic environments cannot help themselves when they offend and should not be blamed and punished. The question this claim raises is why a history of emotional or other deprivation provides an independent ground for a control excuse. Some argue that deprived offenders lack free will, cannot help themselves, or have no choice, but as we have seen, these locutions are usually just conclusions or proxies for more extended arguments that need to be unpacked. Let us address the possibilities.

If conditions of extreme deprivation require offending to save life, say stealing to prevent starvation, then the agent is clearly justified and no control excuse is necessary. Many claim that deprivation can nonculpably cause irrationality, can drive people crazy. If this occurs, however, irrationality and not internal

---

327, 334-39 (1992) (arguing that a diagnosis of multiple personality disorder (“MPD”) may be a result of “artificial production” and represents a “misdirection of effort”). But see Dorothy O. Lewis & Jennifer S. Bard, Multiple Personality and Forensic Issues, 14 PSYCHIATRIC CLINICS N. AM. 741, 755 (1991) (suggesting that MPD is more highly prevalent amongst criminal offenders than commonly thought but is often mistaken for antisocial personality disorders); Elyn R. Saks, Does Multiple Personality Disorder Exist? The Beliefs, the Data, and the Law, 17 INT’L J. L. & PSYCHIATRY 43, 45-70 (1994) (arguing that while the existence of a MPD remains an open question, proponents have a better case than the skeptics). Cf. Seymour L. Halleck, Dissociative Phenomena and the Question of Responsibility, 38 INT’L J. CLINICAL & EXPERIMENTAL HYPNOSIS 298 (1990) (noting that there is considerable controversy about the disorder and recommending that, in the absence of good objective data, clinical wisdom and experience suggest treatment aimed at maximizing the responsibility of those allegedly suffering from MPD).

188 Judge David Bazelon is the most famous modern exemplar of this position. See David L. Bazelon, The Morality of the Criminal Law, 49 S. CAL. L. REV. 385, 403 (1976).

189 The purpose of this Section is not to canvass all possible arguments about the viability of a “rotten social background” excuse, but, rather, simply to determine if control problems are the best justification for such a defense.

190 The explanation for the higher rates of severe mental disorder among the
coercion or a defect of volition or choice excuses. But if deprivation or rotten social background simply produces character flaws or antisocial predispositions, why should this create more ground for excuse than any other nonculpable developmental cause of such flaws or predispositions? Determination or universal causation does not excuse and many people from more advantageous backgrounds turn out rotten nevertheless. Finally, a rotten background that places an agent in environments in which there are fewer opportunities to fly straight does not satisfyingly create an excusing condition unless choice is so nonculpably constrained that coercion occurs. This unfortunate circumstance will rarely arise, however. Even the most impoverished environments in our society provide enough opportunities so that few agents can claim, when the balance of evils is negative, that they meet the moralized criteria for duress or internal coercion. This is not to deny the existence of criminogenic environments that exist as a result of inequality and unjust social laws, institutions, and practices. One might then claim that it is unfair to blame and punish offending agents from those environments. The reason, however, would not be that the agent was excusable. As noted, there is no defect of volition, irrationality, external or internal coercion, or any other standard excusing condition. Rather, the argument would be that such an unjust society lacks the moral authority to blame and punish those who are rightly called its victims.

If deprivation or rotten social background does not satisfy standard criteria for excuse, what is the basis of the powerful intuition that deprived agents should be excused? The true basis, I believe, is sympathy for those who have suffered. Although socially disadvantaged is elusive. The two dominant hypotheses are social causation and social selection or drift. The most recent large-scale epidemiological attempt to resolve the question discovered that social selection may better explain the higher rates of schizophrenia, but that social causation may better explain depression in women and antisocial personality and substance abuse in men. See Bruce P. Dohrenwend et al., Socioeconomic Status and Psychiatric Disorders: The Causation-Selection Issue, 255 Science 946 (1992).
such responses are understandable because sympathy is entirely appropriate, deprivation is nevertheless not clearly relevant to responsibility ascriptions, and the proper legal response to the rational and uncoerced but relevantly deprived criminal is problematic. As Martha Klein has demonstrated, deprivation is arguably relevant only if it is the cause of the actor's *morally reprehensible* state of mind that produced the criminal act.\textsuperscript{194} Klein also suggests that in such cases the offender deserves less punishment because she "has paid something in advance" by her previous suffering.\textsuperscript{195}

The payment-in-advance principle has intuitive appeal, but note that it is not an argument about responsibility. Rather, a fully responsible miscreant is simply being punished less after the offense because she has been punished before the offense by the very conditions that produced her culpability. Viewed from the vantage point of an entire life, the offender has been fully punished, albeit in large measure by agencies other than the state. There are also practical problems with this suggestion, such as identifying causal suffering, calibrating the proper deserved punishment post-offense, and dealing with the danger to public safety that less-punished, responsible actors represent. For our purposes, however, the crucial point is simply that a history of deprivation itself does not furnish grounds for a control excuse, even when it is causally relevant to the actor's offense. If the law wishes to consider the rotten social background of a rational, uncoerced agent, the agent's lack of control is most decidedly not the reason for doing so.

F. The "New Syndrome Excuse Syndrome"

Mental health professionals, lawyers, law professors, and others often use the purported identification of new "syndromes" that may be causally associated with criminal behavior to claim that the syndrome sufferer should be excused because the sufferer could not help offending. Battered victims seeking to defend against homicide or assault charges on an excuse theory, for example, may employ such claims.\textsuperscript{196} The analysis of this Article should demon-

\textsuperscript{194} See MARTHA KLEIN, DETERMINISM, BLAMEWORTHINESS, AND DEPRIVATION 84-91, 172-76 (1990).

\textsuperscript{195} Id. at 82.

\textsuperscript{196} But see Anne M. Coughlin, Excusing Women, 82 CAL. L. REV. 1 (1994) (rejecting
strate, however, that causation, lack of intention, volitional defect, and lack of choice are not generally good candidates to support a control excuse. Once again, let us unpack the possible bases for excuse not to resolve all the issues concerning the defenses available to battered women, but rather for the limited purpose of assessing whether a control theory provides a sound justification for an excuse.

Consider battered victim syndrome sufferers who kill in circumstances that would not satisfy the criteria for a self-defense justification. The killer's conduct is not excusable simply because it is caused. Furthermore, surely most kill intentionally: it is their clear purpose to take the batterer's life. Their volitions execute this more general intention most effectively. If, on the other hand, they kill while in a dissociated state, then the proper analysis is either no action for want of volition, or irrationality.

Claims by battered victims that they had "no choice" are metaphorical and moral, not literal. The battered agent clearly makes a literal choice between at least two options—kill or do not kill—when she takes the batterer's life. The moral claim is ambiguous, however, and at least four possibilities suggest themselves. First, the battered victim might be claiming that there were no genuinely reasonable alternatives to killing the batterer when she did. If correct, the killing is justified and an excuse is not necessary. Second, suppose that a reasonable alternative exists, but the syndrome deprives the sufferer of the ability to recognize either that it does exist or that she can take advantage of it. An excuse might then obtain, but it would be based on irrationality. Third, the killer might claim that the batterer's continued existence caused her such fear, depression, or other untoward emotional states, that she killed to end the unbearable dysphoria. This argument appears to be one

both the "battered woman excuse" because it negatively stereotypes women and the law's theory of responsibility because it fails to accommodate women's experiences); Stephen J. Schulhofer, The Gender Question in Criminal Law, SOC. PHIL. & POL'Y, Spring 1990, at 105, 111-30 (suggesting that the criminal law is "pacific" and should not encourage private violent solutions to interpersonal conflict).


of internal coercion, but once again, such claims seem better analyzed as irrationality problems. Moreover, the dysphoria must be immense to support an internal coercion claim, because the harm caused, homicide, is a great evil. Indeed, the common law would never accept this theory, even if it accepted internal coercion arguments in general, because duress/coercion does not excuse homicides. Last, the battered victim might assert that although alternatives that the law deems reasonable were available, they were in fact insufficient as a matter of justice. The killer should be excused because the law of self-defense immorally left her with no fair choice to defend herself lawfully. Like the argument of the deprived offender, however, this claim is not about an excusable agent. Rather, it is addressed to the morality of the law itself.

In sum, battered victims who strike back should on proper occasions be justified or excused, but not because they were out-of-control agents.

V. ASSESSING AND ADJUDICATING CONTROL EXCUSES

Previous Parts of this Article have argued that control excuses are best understood as irrationality or internal coercion claims, and that the latter often reduce to irrationality claims themselves. This Part examines how to assess claims of lack of control to permit finders of fact to decide whether the agent is responsible.\(^{199}\)

Because the law is concerned with whether the agent was irrational or internally coerced, the fundamental inquiry in all cases concerns the out-of-control agent's psychological phenomenology—what were the agent's thoughts and feelings. Needless to say, we cannot directly "read" each others' minds or measure the strength of desires or feelings. Nevertheless, most people are quite expert at identifying and assessing other agents' reasons for action. Relatively orderly and predictable human interaction is possible only because we are all able within reasonable limits to make inferences about our fellow humans' mental states from behavior, including

speech acts. Moreover, assessing the rationality of another person's reasons for action requires only that we identify those reasons and then evaluate them according to our operative, normative theory of rationality. Of course, how much irrationality is required to justify excusing is a moral and legal matter.

In contrast, judging the strength of another's desires and dysphoria, or fear of it, is a herculean endeavor. Unlike rationality cases, there are no relatively clear phenomena to match against a roughly consensual normative standard. Indeed, this is a major difficulty with the empirical model of internal coercion: famously, we cannot distinguish between irresistible impulses and those impulses simply not resisted. No established metric exists to determine the magnitude of impulses, desires, or feelings. That two independent observers trained in the same system of assessment would agree that a subject exhibits desires of a certain strength or is unable to refrain from acting does not entail that the system is valid, and I know of no such measurement system with established validity. Furthermore, it is difficult to disentangle the strength of desires, the strength of temptations, and the capacity for self-control. There have been numerous studies of impulsiveness and self-control in the psychological and psychiatric literature, and people do commonsensically note individual differences in these traits. Moreover, we talk about impulses, the will, and self-control as if these are independent psychological entities that are well-understood and reliably identifiable. But theoretical disarray abounds in psychology; the studies often contradict each other; measures of supposedly the same variable correlate poorly; findings are often based on suspect self-reports; and, most importantly, the studies do not address, and folk psychology does not know, whether and to what degree people are unable to refrain from acting.

Neither in psychology, philosophy, nor folk psychology is there a reasonably uncontroversial understanding of these matters. Finally,

200 Rogers's system, see supra note 110, which is discussed further in this Section, is primarily about rationality and is unvalidated.


202 See generally THE IMPULSIVE CLIENT, supra note 40. Especially useful for considering the points raised in the text are Dickman, supra note 44, at 151, and McCown & DeSimone, supra note 41, at 3.
we do not know how mental disorder affects self-control in general, apart from its more clear role in affecting perception and belief, which are variables central to rationality.

The strongest contrary claims in the literature fail both conceptually and empirically. For example, in an article about legal insanity that purports to demonstrate that volitional problems can be reliably identified, Rogers provides "Representative Criteria for Assessing Volitional Capacity." But inspection of the criteria Rogers proposes discloses that they are firmly in the camp of folk psychology, and most do not describe failures of volition or the will—which is anyway never defined. Rather, they are criteria of irrationality in the face of strong desires, emotions, impulses, and the like. For instance, Rogers's criteria ask: "What did the defendant perceive as his or her alternatives to the criminal behavior?" or "Did the criminal behavior include planning or preparation?" One criterion begs the question by asking, "Was the loss of control caused by a strong emotional state (for example, rage reaction) or intoxication, or both?" None of these criteria individually nor all of them taken together can demonstrate with acceptable scientific precision whether and to what degree a defendant lacked the capacity to behave lawfully under the circumstances. And virtually all are designed to uncover rationality defects rather than defects of volition. And, in a later article using only four forensic psychiatrists as subjects, Rogers and colleagues claim that they empirically establish that volitional criteria are practically important and logically distinct from cognitive criteria. But the article shows only that the tiny number of subjects believe that they can distinguish and use volitional criteria. No evidence in the study demonstrates that the subjects in fact used volitional criteria that are independent of rationality, and nothing in the study, contrary to its blithe assurance, supports the conceptual validity of independent volitional problems.

Proponents of an independent coercion or volitional excuse often try to justify its adoption in the face of conceptual and assessment problems by correctly arguing that our understanding of the causes of cognitive or rationality defects is as primitive as the

203 Rogers, supra note 110, app. A at 848.
204 Id.
205 Id. (emphasis added).
understanding of the etiology of inner coercion. Although true, this argument is irrelevant to the differential difficulty of assessing existing irrationality and inner coercion. The law's concern is not why glitches occur. Rather, to evaluate responsibility the law needs to know only whether and to what degree glitches do occur. Understanding the causal background may in some cases be probative about whether an excusing condition exists, but no particular cause is required to justify the excusing condition. For example, if we are convinced that a person was in the throes of nonculpable irrationality, we excuse the agent, even if we do not know what produced the abnormality. The causes of cognitive and volitional defects are equally obscure, but for the reasons suggested above, we can empirically identify and normatively assess each others' reasons for action far better than we can identify and assess each others' strength of desire or intensity of feeling. Although there are no conclusive studies that prove this point, I believe that the opposite claim is so counterintuitive that it is fair to place the burden of persuasion on those who disagree.

Ultimately, internal coercion assessment may collapse into rationality assessment. Virtually all cases of so-called irresistible impulse will prove on close analysis to be instances of irrationality, especially if the law continues to require that an abnormality is required. Even the commonsense basis for judging control problems is often a disguised rationality criterion. For example, the "policeman at the elbow" test, which is usually understood as a volitional standard, is, I think, better interpreted as a rationality test. Those who offend in the face of certain capture have either rationally decided for political or other reasons that the offense is worth the punishment, as in cases of civil disobedience, or they are irrational. We generally tend to conclude that intense internal coercion was operative if conduct was so irrational that we cannot make any sense of it; otherwise, why would the person do it? Again, however, rationality is the real issue.

Still assuming, however, that cases of pure internal coercion exist, the best we can do is gather evidence about the defendant's conduct in a wide variety of circumstances that should constrain control difficulties, ask the actor to tell us how she felt, and observe

---

207 One may object that we must identify causes such as mental disorder, but the same evidence that proves the presence of mental disorder also proves the substantive part of an irrationality test. Moreover, there is no need to identify the cause of the mental disorder.
psychophysical signs, such as trembling or perspiring, that may also provide a clue. The moral test asks only for phenomenological description and then weighs it in the moral balance. By comparing the intensity of the threatened dysphoria to the conduct chosen to avoid it, we can make the moral and legal decision whether an internal coercion excuse is warranted. This we can try to do without kidding ourselves by treating the pseudoscientific enterprise of assessing so-called volitional problems as if it were an achievable inquiry based on empirically valid techniques. Even when performed rationally, however, assessments of internal coercion are a dicey proposition at best. On both theoretical and practical grounds, the law should (and does) treat internal coercion claims with great caution.

CONCLUSION: THE MORAL OF THE STORY

The moral of the story is simple: claims about lack of control are far more complicated than most claimants recognize. Often the claim is an ideologically motivated but unanalyzed assertion that this defendant should be excused. On many occasions the defendant should be excused because she was irrational, internally coerced, failed to act, or, more controversially, lacked some other attribute, such as the capacities for empathy and guilt, that make it hard to fly straight and thus should be included in a just account of responsibility. Recognizing the complexity will help lawyers and the legal system fly straight themselves, because it will irresistibly compel them to define in detail, and justify more fully, those criteria that genuinely should excuse.

208 See generally Seymour L. Halleck, Clinical Assessment on the Voluntariness of Behavior, 20 BULL. AM. ACAD. PSYCHIATRY & L. 221, 226-34 (recognizing that clinicians "rarely have access to conceptual or practical guidelines for assessing involuntariness," and failing to provide explicit criteria for involuntariness, but offering "factors" to be considered in assessing involuntariness).