

University of Pennsylvania Carey Law School

## Penn Law: Legal Scholarship Repository

---

Faculty Scholarship at Penn Law

---

2002

### Excuses and Dispositions in Criminal Law

Claire Oakes Finkelstein

*University of Pennsylvania Carey Law School*

Follow this and additional works at: [https://scholarship.law.upenn.edu/faculty\\_scholarship](https://scholarship.law.upenn.edu/faculty_scholarship)



Part of the [Criminal Law Commons](#), [Criminal Procedure Commons](#), [Criminology Commons](#), [Criminology and Criminal Justice Commons](#), [Ethics and Political Philosophy Commons](#), [Jurisprudence Commons](#), [Law and Society Commons](#), and the [Social Control, Law, Crime, and Deviance Commons](#)

---

#### Repository Citation

Finkelstein, Claire Oakes, "Excuses and Dispositions in Criminal Law" (2002). *Faculty Scholarship at Penn Law*. 1003.

[https://scholarship.law.upenn.edu/faculty\\_scholarship/1003](https://scholarship.law.upenn.edu/faculty_scholarship/1003)

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact [PennlawIR@law.upenn.edu](mailto:PennlawIR@law.upenn.edu).

# Excuses and Dispositions in Criminal Law

Claire Finkelstein†

## I. INTRODUCTION

What do criminal laws prohibit? A series of immoral or harmful acts? Or does the law also seek to prohibit performing those acts for certain reasons, acting on certain motives, or acting on the basis of certain character traits? In short, does the law look only at the quality of the act the defendant performs? Or does it look more broadly at whether the person who performed the prohibited act was righteous or ignoble, well-meaning or malign?

The traditional view says that the criminal law focuses uniquely on acts. According to that view, character and motive are irrelevant to criminal liability.<sup>1</sup> They are not relevant to the *actus reus*, since they merely serve to identify the prohibited conduct.<sup>2</sup> And they are not relevant

---

† Professor of Law and Philosophy, University of Pennsylvania. My thanks to Kim Ferzan, Leo Katz, and Peter Westen for comments on various drafts and discussions on the topic of this article.

1. See H.L.A. Hart, *Punishment and Responsibility* (1968); R.A. Duff, *Virtue, Vice and Criminal Liability: Do We Want An Aristotelian Criminal Law?*, 6 *Buff. Crim. L. Rev.* 147 (2002); Benjamin Sendor, *The Relevance of Conduct and Character to Guilt and Punishment*, 10 *Notre Dame J.L. Ethics & Pub. Pol'y* 99 (1996); R.A. Duff, *Choice, Character and Criminal Liability*, 12 *Law & Phil.* 345 (1993); Michael S. Moore, *Choice, Character, and Excuse*, 7 *Soc. Phil. & Pol'y* 29 (1990).

2. Motive is occasionally relevant for establishing the special mental state required for crimes of specific intent, as it is in bias crimes or crimes that separately criminalize killing for hire or for sexual pleasure. But even the ordinary specific intent crimes, such as burglary, which requires the "intent to commit a crime," Model Penal Code § 221.1 (1) (Proposed Official Draft 1962), or forgery, which requires a "purpose to defraud or injure," *id.* § 224.1(1), do not actually criminalize motive, though they might appear to do so. The *mens rea* in these cases does not so much go to the question of why the defendant entered a building or altered a writing, as it goes to the *further* intention with which he did it. The motive for such crimes, such as money, revenge, or hatred, are nowhere part of the offense definition.

German law, by contrast, explicitly treats crimes done from different motives as different crimes. For example, the StGB defines murder in terms of its motive, namely as killing for sheer pleasure, to satisfy a sexual impulse, or on

to the mens rea, since the mental state requirement serves only to ensure that the defendant is responsible for what he did or for the state of affairs he brought about.<sup>3</sup> The traditionalist will of course allow that there are instances in which character or motive can affect a defendant's ultimate fate. A motive like self-defense or necessity will entitle a defendant to a justification defense. And a person who has manifested a good character throughout his life will usually receive a shorter sentence than a defendant who has manifested a bad one. But these examples are not inconsistent with the traditional view, which maintains that notions like character and motive are irrelevant for *prima facie* judgments of criminal responsibility.

In recent years, a number of writers have challenged the received wisdom.<sup>4</sup> Inspired at least in part by a revival of virtue ethics in the philosophical literature,<sup>5</sup> the "character theorists" begin with the Aristotelian thesis that human beings create their own character traits by repeatedly performing acts of the sort that display that trait. Thus a person acquires the trait of bravery by performing frequent brave acts, and a person acquires the trait of liberality by spending money freely. According to the character theorists, punishing agents for bad acts they perform is a way of punishing them for the bad characters

---

account of greed or other "base motives." But ordinary intentional killing, not done from one of these motives, is not murder. §§ 211 & 212 StGB.

3. A person does not lack the mens rea for murder, for example, because he killed for the purpose of putting a terminally ill patient out of his misery, or even because he killed in self-defense. That is, justifications like self-defense and necessity do not "negative" the mens rea for the offense. A person who kills in self-defense kills intentionally, and so satisfies the mens rea for murder. This is in contrast to excuses like mistake, which negative the mens rea for the crime they exonerate.

4. See Victor Tadros, *The Characters of Excuse*, 21 *Oxford J. Leg. Stud.* 495 (2001); Kyron Huigens, *Virtue and Inculpation*, 108 *Harv. L. Rev.* 1423 (1995); Dan M. Kahan & Martha Nussbaum, *Two Conceptions of Emotion in Criminal Law*, 96 *Colum. L. Rev.* 269 (1996); John Gardner, *The Gist of Excuses*, 1 *Buff. Crim. L. Rev.* 575 (1998); Peter Arenella, *Character, Choice and Moral Agency*, 7 *Soc. Phil. & Pol'y* 59 (1990).

5. See Philippa Foot, *Morality as a System of Hypothetical Imperatives*, 81 *Phil. Rev.* 305 (1972).

that produced those acts, and of encouraging the development of good character traits in the future.

The debate between the traditionalists—the so-called “choice theorists”—and the character theorists has particularly come to the fore in writings about the criminal defenses. Because choice theorists see human beings as responsible for the freely chosen conduct in which they engage, choice theorists are inclined to understand criminal defenses as falling into one of two categories: Action which is chosen but justified by considerations of social welfare, and action which is excused because not fully chosen.<sup>6</sup> But character theory has allowed for an interesting and different position on the defenses. If criminal liability is based on the bad character or motive of the person who intentionally violates a prohibitory norm, a defendant might at least be entitled to an excuse where his bad act does not in fact support an inference from the quality of the act to the quality of the underlying character trait.<sup>7</sup> Because we cannot draw the conclusions about the defendant’s character we supposedly do in cases of unexcused wrongdoing, the rationale for inflicting punishment does not obtain.<sup>8</sup>

A variation on this general theme suggests that the criminal law is primarily concerned with punishing defendants who manifest a defect of motivation, and that we should excuse those who violate the law in cases where no defect of motivation exists.<sup>9</sup> Like the claim about

---

6. An alternative approach to justification in the choice theory literature sees justified acts as issuing from moral rights people have. As I explain below, however, the rights version of justification does not fundamentally change the structure of the choice theory, and in particular, does not eliminate its drawbacks. See *infra* text accompanying notes 34.

7. George Fletcher, *Rethinking Criminal Law* § 10.3.1 (1978); Huigens, *supra* note 4.

8. See Michael Moore, *Placing Blame* (1997); Stephen Garvey, “As the Gentle Rain From Heaven”: Mercy in Capital Sentencing, 81 *Cornell L. Rev.* 989, 1024 (1996) (referring to what he calls the “character-will theory”); Michael Corrado, Notes on the Structure of A Theory of Excuses, 82 *J. Crim. L. & Criminology* 465 (1992).

9. See Richard B. Brandt, A Motivational Theory of Excuses in the Criminal Law, in *Nomos XXVII: Criminal Justice* 165 (J. Roland Pennock & John W.

character, the suggestion about motive is that we cannot draw any negative conclusions about the motives of a person whose reason for violating the law is that he fears for his life or for the life of a loved one.

The purpose of this article is not to revisit the debate between the choice and the character theorists. Much has already been written on that debate, and I have little of substance to add to it. Instead, my purpose is to return to a project I began in previous writings, namely to understand the structure of a certain class of defenses I call "rational excuses."<sup>10</sup> Defenses in this category straddle the line between traditional excuses and justifications. They are cases, I claim, where the agent acts on the basis of deliberation, and so controls her behavior. But I call them "excuses," rather than justifications, because they are also cases where the agent acts for personal reasons, rather than for the impersonal considerations of social welfare. My suggestion will be that neither the choice theory nor the character theory is particularly well-suited to account for the rational excuses. In particular, I shall argue that while choice theorists are more nearly correct about the general structure of criminal responsibility, they do not have a plausible theory of exculpation. And while character theorists are right to suggest that states of character and dispositions should be taken into account in criminal liability, the theory does not do so in the right way. The rational excuses thus reveal the need for a different theory of exculpation than either the choice or the character theory provides.

In what follows, I shall suggest that the rational excuses exonerate because they are cases in which the agent acts on the basis of what I shall call an "adaptive disposition," namely a disposition that enhances the welfare of the agent who cultivates it and makes possible

---

Chapman eds., 1985), rpt. in Richard B. Brandt, *Morality, Utilitarianism, and Rights* (1992).

10. See Claire Finkelstein, *Self-Defense as a Rational Excuse*, 57 *U. Pitt. L. Rev.* 621 (1996); Claire Finkelstein, *Duress: A Philosophical Account of the Defense in Law*, 37 *Ariz. L. Rev.* 252 (1995).

collective welfare improvements. The law is interested in promoting and rewarding such dispositions, because there will be mutual gains from their adoption if members of society generally possess them. It will thus turn out to be possible to vindicate the legal rules establishing the class of defenses I am calling rational excuses in contractarian terms, since a rational agent would agree to a legal rule that leaves him better off than he would be in its absence.

My argument will proceed as follows. In the next Part, I review the basic elements of the choice and character theories, along with the account of the defenses implied by each. In Part III, I lay out the structure of the rational excuses, and I explain why the choice theory is ill-equipped to account for defenses in this category. In Part IV, I do the same for the character theory. I argue that that theory is both problematic in its own right and unable to account for the rational excuses. In Part V, I present my alternative account of the rational excuses. I explain why the rational excuses require a disposition-based account. In Part VI I conclude by explaining why the fact that the relevant dispositions are adaptive allows us to justify the rules governing the rational excuses in contractarian terms.

## II. CHOICE AND CHARACTER

The choice theory of criminal responsibility predicates responsibility on the individual choice the defendant makes to violate a prohibitory norm. The choice theory thus naturally focuses on actions, rather than on states of character, and on whether those actions conform to an objective standard of conduct. As Peter Arenella explains the theory, "When . . . an actor makes a rational choice to engage in conduct that breaches the governing norm, he deserves moral blame because he could have chosen to comply with it."<sup>11</sup> On other versions of the account, what is relevant is the moral worth of the choice the agent makes. As R.A. Duff writes, "We might dislike or regret other

---

11. Arenella, *supra* note 4, at 59.

aspects of a person's character, but only his will is a proper object of moral criticism."<sup>12</sup> It is consistent with this view to treat each criminal act as a separate object of evaluation. The choice theory has no interest in a pattern of choices an individual might have made in the past, nor in an overall trait that might lead her to make such choices in the future. For this reason, the choice theory has no use for observations about the relation between the prohibited action and the defendant's character.

The authoritative statement of the choice theory is H.L.A. Hart's in *Punishment and Responsibility*. On the conception of responsibility for which Hart argues, the crucial questions, both for moral and legal responsibility, are about "the character or extent of a man's control over his own conduct, or . . . the causal or other connexion between his action and harmful occurrences, or . . . his relationship with the person who actually did the harm."<sup>13</sup> The basic element of control must also be combined with a set of background capacities to obey the law, as well as with external circumstances that allow the agent to exercise those capacities: "What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise those capacities."<sup>14</sup> When the background capacities are satisfied, and the external circumstances provide individuals with an opportunity to obey the law, the choice to perform an (unjustified) act that violates the law is both a necessary and a sufficient condition for criminal liability.

The choice theory is usually combined with a welfare-based standard for evaluating the content of an agent's choices. Thus an agent will be blamed for his decision to violate the criminal law, unless he can justify his choice by its positive effects on social welfare. Such justifications apply to instances in which, although the defendant's act

---

12. Duff, *Choices, Character and Criminal Liability*, supra note 1, at 346.

13. Hart, supra note 1, at 225.

14. *Id.* at 152.

falls within the ambit of the prohibitory norm the law defines, the law is otherwise committed to promoting the defendant's conduct. A person who must burn a field in order to create a firebreak to save a town from destruction by forest fire would have a defense to the crime of intentionally destroying the property of another, because the overwhelming social importance of saving the town justifies the legal infraction. For those criminal acts that do not merit punishment, but where reasons of social utility are unavailing, the choice theory can only explain the defendant's exoneration if it is able to deny that the agent actually *chose* to violate the prohibitory norm. Where the agent was mistaken, acted accidentally, or moved his body involuntarily, we cannot regard him as having *chosen* to violate the law, and thus we do not hold him responsible for his conduct. Lack of capacity in the criminal law thus functions as an excusing condition much in the way that a denial of capacity will serve to invalidate a contract: Just as a person cannot be understood to have truly agreed to the terms of a contract if he was insane, infantile, intoxicated, mistaken, coerced, etc., so these same conditions will defeat the presumption that an agent who violated a criminal prohibition was responsible for violating the law. In this way, the choice theory gives rise to what is sometimes called the "capacity" theory of the excuses, meaning that it accounts for the excuses in terms of a diminished capacity on the agent's part to conform her behavior to the law.<sup>15</sup>

The association between Hart's choice theory of responsibility and the capacity theory of the excuses has become a standard part of commentary and black letter law. Paul Robinson, in his treatise on the defenses, explains that all excuses are characterized by the fact that they have a "disability requirement."<sup>16</sup> He writes: "The disability requirement is the actor's abnormal condition at the time of the offense . . . . It may be a long-term or even

---

15. See Tadros, *supra* note 4, at 495.

16. 2 Paul H. Robinson, *Criminal Law Defenses* § 161 (1984).



permanent condition, such as subnormality, or it may be a temporary state, such as intoxication, somnambulism, automatism, or hypnotism. Its cause may be internal, as in insanity, or external, as in duress."<sup>17</sup> A defendant will normally have a defense, Robinson suggests, when his disability causes him to act in a way that is not the product of his voluntary effort or determination, to mistake the physical nature or consequences of his conduct, to mistake the wrongfulness or criminality of his behavior, or to be unable to control his conduct. If, on the other hand, the defendant chose to violate a criminal prohibition, while he was in control of his conduct and without legal justification, he is responsible for the violation and will be called to account for it.

Those who take a character-based approach to criminal liability, by contrast, regard the focus on the defendant's choice as an overly narrow basis on which to assess the merits of his conduct. Instead, they argue, criminal acts are punishable only insofar as they stem from established features of an agent's character. By "character," these theorists have in mind roughly the Aristotelian concept of fixed or stable aspects of an agent's psychological make-up that pertain to virtues and vices. Unlike their counterparts in moral theory, however, the character theorists in criminal law need not hold a virtue theory of moral responsibility. The relevance of character to criminal liability has been claimed both by those who hold a retributivist theory of punishment, and by those who espouse the corresponding utilitarian view.<sup>18</sup> The suggestion of the former (what we might call "character retributivists") is that an actor's desert for punishment is

---

17. *Id.* at 222.

18. As Michael Corrado explains:

A character theory . . . does not have to be a utilitarian theory. A retributive theory of excuses . . . might assert that desert is the only appropriate basis for punishment. Only bad character deserves punishment, on such a theory, and we excuse cases in which the agent could not have done otherwise because the inability to do otherwise prevents the inference from the illegal act to the bad character.

Corrado, *supra* note 8, at 470.

based on conclusions his act allows us to draw about the bad state of his character. George Fletcher, for example, argues that "[a]n inference from the wrongful act to the actor's character is essential to a retributive theory of punishment."<sup>19</sup> The claim of the latter ("character utilitarians") is that the criminal act allows for an inference to bad motivation or character, and that a system of punishment that penalizes on the basis of such features will maximize social welfare.<sup>20</sup>

The integration of character into either a retributive or a utilitarian theory of punishment produces a character-based theory of the excuses: If punishment is for bad states of character, then an excuse is warranted when an agent who has violated a criminal norm does not manifest the character defect we would normally infer from the act he performs. Thus a character retributivist would claim that a defendant who robs a bank under duress does not deserve to be punished. As Fletcher writes, "[I]f a bank teller opens a safe and turns money over to a stranger, we can infer that he is dishonest. But if he does all this at gunpoint, we cannot infer anything one way or the other about his honesty."<sup>21</sup> A character utilitarian would say that there is utility only in punishing people who manifest bad character.<sup>22</sup> On Richard Brandt's version of character utilitarianism, the thesis is spelled out in terms of motivation, rather than character.<sup>23</sup> On this view, the purpose of the criminal law is to punish for defects of motivation; a worse defect of motivation should entail a more severe punishment.<sup>24</sup> The attending theory of the

---

19. Fletcher, *supra* note 7, at 800.

20. Brandt, *supra* note 9.

21. Fletcher, *supra* note 7, at 800.

22. See Paul H. Robinson & John Darley, *The Utility of Desert*, 91 *Nw. U. L. Rev.* 453-499 (1997).

23. Brandt himself seems to think the two equivalent, since he thinks of character as a "system of motivations." Brandt, *supra* note 9, at 171.

24. Brandt writes: "if a worse defect of motivation is shown by intentional homicide than by reckless homicide, one might infer that the punishment for murder should be more severe than for manslaughter." Brandt, *supra* note 9, at 170.

excuses would exempt an agent who commits an illegal act when that person did not manifest a defect of motivation: "a legally wrongful act ought to be subject to punishment only if the act manifested a defect of motivation."<sup>25</sup>

The difference between the character and the choice theory can be captured in terms of two kinds of evaluations or judgments a person might make.<sup>26</sup> On the one hand, there is evaluation of a person in light of something he did, where we evaluate the person's action in isolation from facts about him that might explain why he behaved as he did. This form of evaluation is focused on action because it is concerned with whether the agent's behavior conformed to an external norm of permissibility. On the other hand, there is evaluation of an agent as a person, taken as a whole, where we are not particularly interested in whether his behavior conforms to a norm, but in whether he is a good or bad person, considered in a more general sort of way. The choice theorists think of criminal responsibility as a form of the first kind of evaluation, whereas the character theorists appear to think of it as a form of the second kind.

While I am not prepared to defend the choice theorist's approach to criminal liability across the board, I do believe it has this over the character theory: the character theory appears to efface the distinction between the two kinds of evaluation outlined above. It confuses criminal responsibility with the kind of judgment that might interest a priest, a psychotherapist, or a governor deciding whether to grant clemency. The governor, for example, is truly concerned with the state of the criminal's soul. He wishes to know whether a terrible deed the criminal has performed is out of keeping with his overall moral worth, and so whether the criminal is a fitting object for an exercise of mercy. In order to decide whether the deed is in fact representative of the criminal as a person, the

---

25. *Id.*

26. For a slightly more extended discussion of the distinction, see Claire Finkelstein, *The Irrelevance of the Intended to Prima Facie Culpability: Reply to Moore*, 76 *B.U. L. Rev.* 335, 342-344 (1996).

governor wants to know what motivated the conduct, and whether the criminal has performed similar acts in the past. He is also interested in facts that have nothing to do with the wrongful act itself, such as whether the criminal would be a productive member of his community if returned to normal life, and whether others in society would benefit or be harmed by his release. Unlike a judge or members of a jury who are concerned with guilt or innocence, the governor's interest in the criminal's act is limited to the *evidence* it provides of that person's internal moral worth and of his future dangerousness. He has no more reason to be interested in what the criminal did on this occasion than he would be in any other manifestation of character or motivation.<sup>27</sup>

Despite the fact that the choice theory seems to better capture the basic form of evaluation in criminal law, it faces some important difficulties. In the next Part we will focus on the rational excuses, for it is here that the choice theory's shortcomings are most apparent. We will see first, that it is difficult to make sense of the rational excuses in the absence of the notion of a disposition, and second, that the choice theory's difficulty explaining responsibility for acts that arise "habitually" is what makes it unsuited to account for this class of defenses.

---

27. Michael Corrado argues that the choice and character theories reduce to the same thing. The character theory claims that it is unfair to hold agents liable for acts that are out of character, but fair to hold them for acts in which their settled dispositions are expressed. And as he understands the choice theory, it maintains that it is unfair to hold agents liable for actions caused or constrained by external forces, and fair to hold them for actions they themselves author. Since action that arises from a disposition or character trait is "internally" produced, he thinks both theories would hold agents responsible for bad acts that stem from the bad character traits they author. Michael Corrado, *supra* note 8, at 479-80. I think, however, that Corrado misunderstands the choice theory in this context. An action that stems from a disposition is not separately willed in the sense the choice theorists have in mind, since it is not chosen in its own right. Rather, an action that stems from a character trait is a non-optional result of the earlier choice or choices that led to the development of the character trait in the first place. So although actions that stem from a disposition may be determined by something "inside" the agent, they are not chosen in the way the choice theorist requires.

## III. RATIONAL EXCUSES AND THE CHOICE THEORY

Consider the following instances where a defendant might claim partial or total exoneration. There is the case of the two sailors who are survivors of a shipwreck, trying to stay afloat by clinging to a plank in the water. The plank is strong enough to support only one, and the defendant saves his own life by pushing his fellow off it. There is the case of the person who commits a criminal act upon demand, after being threatened with the death of his loved ones by members of a criminal organization. There is the person who throws a hand grenade at a mob of young children who are wielding deadly weapons, where he believes this is the only way he can save his life. Or the case of the person enticed by a police officer into buying drugs, where the officer exerts intense pressure to induce the defendant to make the purchase. Finally, there is the person who meets his loved one's rapist unexpectedly, and attacks him in revenge, or the man who comes home to find his wife in bed with another man, and in a fit of rage, shoots them both dead.

My suggestion is that the above cases have a common structure, despite their historically different treatment in the law. On the one hand, they are all cases in which the defendant can be held to account for his conduct, since he is fully in control of the criminal act. Indeed, there is a firmer basis for ascribing responsibility in such cases than control, since the defendant even performs the prohibited act intentionally, and not as a mere side effect of his behavior. But they are also cases in which the conduct is not justified by considerations of the greater social good. That is, the defendant cannot vindicate his conduct by claiming that he does more good than harm. Indeed, his reason for doing what he does in such cases is always personal. The value he perceives in the criminal act stems from the fact that it advances his interests or the interests of someone he cares about. Defenses of the above sort thus appear to hover between justifications and excuses, at least as these categories have been traditionally conceived.

What does the choice theorist do, then, when faced with these examples? The answer is that he squeezes them into the surrounding categories of social benefit, on the one hand, and diminished capacity, on the other, calling the first set of defenses "justifications" and the second set "excuses." The case of the men clinging to the plank, for example, might be treated as an instance of the justification of necessity, and the person who kills the children with the hand grenade could claim the justification of self-defense. The traditional account would explain this by assimilating such cases to more clear-cut examples of necessity and self-defense: the person who burns down a field to create a firebreak in order to prevent a town from being engulfed by flames; the person who attacks a malicious attacker in order to save himself from a lethal attack. On the other hand, the choice theorist would probably treat the person coerced into assisting in the performance of the criminal act and the person pressured into purchasing drugs as excused, according them the defenses of duress and entrapment respectively. The person who meets his loved one's assailant is thought entitled only to the partial excuse of extreme emotional disturbance or provocation.<sup>28</sup> These cases are treated as excuses by assimilating them to cases in which the defendant's conduct is involuntary or otherwise lacking in control. The Model Penal Code embodies the capacity theory perspective when it requires that the coercion be such "that a person of reasonable firmness in his situation would have been unable to resist."<sup>29</sup> Robinson, for example, writes that "the excusing condition in duress is the impairment of the actor's ability to control his conduct."<sup>30</sup> And he argues that entrapment is like the duress defense,

---

28. Under the traditional rules governing the provocation defense, merely encountering a loved one's assailant would be insufficient to generate a defense. The defendant must actually have witnessed the assault immediately prior to his own response. See, e.g., *State v. Shane*, 590 N.E.2d 272 (Ohio 1992) (rejecting claim of provocation for defendant who learned verbally of fiancée's infidelity).

29. Model Penal Code § 2.09 (1) (Proposed Official Draft 1962).

30. 2 Robinson, *supra* note 16, § 177 (b)(1), at 351.

insofar as it is based on the idea that "the defendant's actions were not fully his own."<sup>31</sup> The choice theorist would see the person who merits a provocation defense as someone who lacks the capacity to restrain his emotions. He would prefer the psychological formulation of the defense to the traditional normative formulation. One would thus expect him to side with the Model Penal Code's extreme emotional disturbance formulation over the traditional provocation or "heat of passion" approach.<sup>32</sup>

Evidence that something is amiss with the choice theory's approach, however, comes with the first group of examples we considered, the cases that would traditionally be called justifications. The problem with treating the plank case and the defense-against-children case as justifications is that they are all instances in which the criminal act is not particularly redeemed by considerations of social welfare. Is it truly better, from the standpoint of social utility, that one stranded sailor survive rather than another? The problem is that where each person is concerned, it is not better for the world that *he* survive, and so we cannot say, as between the two, which one would be entitled to push the other off the plank. That is, if the entitlement a person sometimes has to violate the law is based on the social utility of his doing so, then arguably we do not have a reason in a case like the plank case for siding with one man clinging to the plank over the other. And so we do not have a basis for finding one man justified in pushing the other off, that would not also lead us to conclude that the second man is justified in pushing the first man off.

Admittedly, it is better that one stranded sailor survive than that *none* survive, and from this perspective the law ought to encourage someone to push someone else

---

31. *Id.* § 209 (b), at 513. He does allow, however, that there is a different theory of the defense, based on the idea that it is a defense designed to deter improper police conduct.

32. Cf. the "extreme emotional disturbance" defense of the Model Penal Code, § 210.3 (Proposed Official Draft 1962), and a traditional "heat of passion defense," such as Cal. Pen. Code § 192(a) (West 2002).

off. Why not say, then, that *each* is justified in pushing the other off the plank, since society would be better off if *either* one did so? Since the best social rule would probably allow each to try to push the other one off, what impediment would there be to allowing the victor in such a contest to claim the benefit of justification once he succeeds? The problem with this solution has to do with the traditional approach to justification. According to the traditional logic, if A is justified in pushing B off the plank, it would follow that B is not justified in resisting A. That is, the traditional approach denies that each of two individuals can be simultaneously justified, if the actions they perform entitle each to oppose the other. From this it follows that neither man clinging to the plank could be justified, despite the fact that it might be socially beneficial if one were to push the other off.<sup>33</sup>

One might, however, suppose such difficulties peculiar to the plank case, where the claimed entitlements over a common resource are exactly symmetrical. And if this were the case, we need not worry, since where cases of conflicting claims of justification are concerned, precise symmetry is rare. But in fact the problem is common to all cases in the in between category of defenses we are considering. Notice, for example, that we cannot explain the entitlement of the defendant to use lethal self-defense against the children in terms of social utility, since it is not particularly welfare-maximizing for a large number of innocent children to die instead of for a single person attacked by those children to die. Indeed, from the standpoint of social utility, we should favor the children's lives over the victim's, since the former are more numerous. If the theorist is going to treat all self-defense cases as instances of justification, then, he will have to conclude that a person is not entitled to defend himself against a mob of infantile aggressors, and that is surely the wrong result.

---

33. For a further discussion of this case, see Claire Finkelstein, *Two Men and a Plank*, 7 *Legal Theory* 279 (2001).



On the social welfare approach to justification, that an act would increase the level of social utility provides a basis for allocating legal rights. But in the face of the argument just given, one might want to reject the social welfare view of justification, and argue that the legal rights that constitute justifications are identified by an underlying theory of moral rights instead. On this alternate view of justification, it would not be necessary for the prohibited act to be welfare-maximizing in order for it to be legally justified. We would say that the person who violates the law in such cases is justified, because he is defending an antecedent moral right he has. Thus the person attacked by the blameless children would be entitled to defend himself against their aggression on the grounds that he had a moral right to do so, not on the grounds that he would thereby increase social welfare.

But turning to a rights-based approach will not help with the above difficulty. For we still confront the problem of conflicting justifications we saw with the plank case. If A has a moral right to push B off the plank, B must have a symmetrical moral right to do the same to A. But if A is *justified* in pushing B off, B may not rightfully oppose him. Surely if B tries to push A off the plank, he is opposing A in his efforts to push B off. The same might be said for self-defense. If the victim has a right to kill the child-attackers with a hand grenade, then surely the children must also have a right to defend themselves by, let us imagine, using a shield to lob the hand grenade back to the victim. But if this is so, then the victim's moral right to defend himself cannot provide the basis for a legal right of justification, since if the victim is justified in throwing the grenade, the children cannot be justified in using a shield to lob it back.<sup>34</sup> So turning from a social welfare to a rights-based approach

---

34. The situation is actually more complicated than this, since on many accounts of the nature of moral rights, the victim could not have a *moral* right that could be rightfully opposed either. For a discussion of such cases and the difficulties that surround them, see Claire Finkelstein, *On the Obligation of the State to Extend a Right of Self-Defense to its Citizens*, 147 U. Pa. L. Rev. 1361 (1999).

to justification does not seem to make it easier to fit these cases into a justification framework.

Now consider the two cases the choice theory would traditionally place under the heading "excuse." The classification there is equally problematic. For example, in order to make duress cases consistent with the choice theory's approach to excuses, the theory must maintain that a defendant who acts under duress has lost control over his conduct. But what reason is there to suppose that this is the case? The person who decides to rob a bank under threat that his loved ones will be killed if he does not is not less in control of his conduct than the man who pushes the other person off the plank, or the person who attacks a mob of child assailants in self-defense. All three defendants act on the basis of reasons, and the reasons, in both cases, are roughly the same: the agents wish to protect themselves or the lives of those they love. It is often said that the defendant in a duress situation loses control because of the pressure of the situation in which he finds himself. But surely a person who is on the verge of drowning or a person who is about to be killed by a mob of children would be under at least as much psychological pressure.

The same difficulty arises with the choice theory's approach to a person who violates the law under pressure from a law enforcement officer. What reason is there to believe that the defendant in such cases is not in control of his actions or that his conduct has somehow become involuntary? As Robinson himself admits, the entrapment defense cannot be consistently formulated in terms of lack of capacity, since if the person entrapped truly lacks the ability to conform his conduct to the law, the defendant who is induced by *non-government* agents to violate the law should also have a defense.<sup>35</sup>

---

35. Robinson somewhat equivocally concludes that the entrapment defense "probably reflects a combination of concerns including an estoppel notion that it is unfair to permit the entity that has entrapped to then punish." 2 Robinson, *supra* note 16, § 2.09 (b), at 516.

The partial defense of provocation comes closer to fitting the capacity model, since such cases at least involve an altered psychological condition. But it is not quite accurate to think of emotional upset as a loss in a person's capacity to control his conduct. To see this, notice that if a person who encounters his daughter's rapist at a social gathering is excused because he lacks the ability to conform his conduct to the law, then someone who encounters his daughter's ex-fiancé should also have a partial defense if his anger over the broken engagement is what led him to attack the victim. Or if the person who finds his wife in bed with another man receives a defense for killing her and/or her paramour because he is unable to control his rage, then the person who finds his ex girlfriend in bed with her husband in their own home should also have a partial defense, assuming he too kills out of jealousy. Yet the law does not hand out partial defenses to people who kill simply because they lose their tempers. And that suggests that lack of capacity is not the reason we exonerate in such cases. Instead, the law offers a defense in situations in which any *reasonable* person would have done the same. The defense is thus at least partially based on a normative stance we take towards people's emotional reactions, and not primarily on a judgment that they are no longer responsible agents.

If fully responsible defendants who choose to violate the law are to be afforded a defense, it must be because the law endorses the *content* of the reasons on which they act. But it is not clear in any of these cases why the law would endorse the agent's reason for acting, since there is no reason for the law to favor the illegal act over the legal rule it contravenes. Nor is there any reason to suppose that the defendant's behavior is endorsed on a rights-based theory of justification, since these are cases in which it is problematic to think of any moral right the defendant might have as giving rise to a legal right to violate the law. In the face of these difficulties, the choice theory retreats from the traditional claims about social utility to a psychological claim about diminished capacity. But there

is no basis for thinking of most of the above defendants as lacking in capacity either. In particular, the defendants who would have "excuses" on this view are typically as much in control of their conduct as the defendants who would normally have "justifications." There is thus reason to think that the explanation for defenses of the above sort has eluded the traditional theorists, and that some rethinking of the grounds for such defenses is in order.

I have set forth the difficulties with the choice theory's approach to the rational excuses in greater detail elsewhere.<sup>36</sup> I shall not, therefore, elaborate on that theory further here. Let us then turn to the character theory instead.

#### IV. RATIONAL EXCUSES AND THE CHARACTER THEORY

The character theory holds out hope of a better account of the rational excuses, insofar as it is sensitive to the underlying dispositions from which criminal acts can arise. Most helpfully, the character theory has no need to treat the excuses in terms of diminished capacity, since it allows that we can excuse conduct performed by responsible agents. It thus contemplates a normative basis for excusing defendants under certain circumstances. This is attractive, first, because it provides a less distorting account of some of the defenses with which we are concerned, and second, because it holds out the hope of accounting for the defenses we are calling rational excuses in a unified way.

Perhaps the most compelling case that the character theory can explain the rational excuses can be made with respect to the defense of entrapment. There courts are explicit, under at least one line of cases, that the defendant merits a defense when pressured by law enforcement agents into breaking the law, unless the prosecution can

---

36. Finkelstein, *Duress*, *supra* note 10. I did not label the view I was criticizing as such in that article. But insofar as my criticisms were directed towards the traditional approach to the excuses, the choice theory was my implicit target.

show that the defendant was "predisposed" to commit the legal violation anyway. The cases suggest that if a defendant possessed the relevant disposition, the fact that a law enforcement officer pressured him into committing a crime should not provide a defense. A defendant will thus be entitled to a defense if his character is good, meaning that he is not normally disposed to violate the law, and he will lack a defense if his character is bad, meaning that he is inclined to violate the law. Arguably, this same analysis can be extended to cases of duress, as well as to some self-defense and personal necessity cases. While the case law does not explicitly bear out the emphasis on character in the way that the entrapment cases do, the character theorist might nevertheless maintain that the reason we extend these defenses where we do is that they apply to situations in which we cannot infer anything negative about the character of the defendant who violates the law, despite the fact that a criminal violation would normally allow us to draw that inference.

Despite its superficial appeal, the character theory turns out to possess a series of flaws that make it ultimately unable to provide an account of criminal responsibility generally, and of the rational excuses in particular. I shall begin by considering several objections to the character theory of a general sort, and then move on to its specific shortcomings with respect to the rational excuses.

First, the law simply does not seem terribly interested in character when it makes determinations of guilt and innocence, or even when it determines the appropriate sentence for a person found guilty. In particular, having a bad character is neither necessary nor sufficient for justified punishment. It is not necessary, because a person with a good character can commit a terrible crime, and while the evidence of his good character may serve to influence his sentence within a certain limited range, the effect of a good character on the sentence of a person who has committed a very serious infraction is marginal. It is not sufficient, because no matter how bad a person's

character, he cannot be punished unless he has violated one of the criminal law's prohibitory norms. People who are chronically ungenerous, or who regularly deceive others about small matters in order to better themselves, probably have bad characters, or at least bad character traits. But we do not normally think they deserve to suffer punishment for such traits. The motivational version of the character theory is subject to the same criticism: it is neither necessary nor sufficient that a person have a bad motive to warrant criminal punishment. It is not necessary, because the person who holds up a liquor store cannot defend himself by pointing out that he intended to give the money to the March of Dimes. And it is not sufficient, because people may do all sorts of things with the hope of harming their fellows that the law does not attempt to prohibit. The problem, from the character and motivation theorists' perspective, then, is that neither is able to offer a clear principle that accurately distinguishes those immoral or ill-motivated deeds that are criminalized from those that are not.<sup>37</sup> Whatever principle the character theorist might present, he surely cannot avoid the conclusion that we do not criminalize bad character *per se*.

Second is an evidentiary point: Character theorists depend on the claim that where the excuses do not apply, we can infer something about the character of a person from the criminal act he performs. But there seems no reason to suppose this is correct. In some contexts, the fact that a person was willing to violate the law says something *good* about his character—it might suggest, for example, that he is so driven by his moral convictions that he is willing to expose himself to a risk of punishment in order to defend them. This is surely the case when the defendant takes himself to be violating an unjust law or when he is pursuing charitable ends. Robin Hood provides the most famous example, but there is also the savvy detective, the street-wise cop, or the activist lawyer, all of whom play a

---

37. See Leo Katz's contribution to the present volume for many more examples of immoral agents, most of whom could be said to have bad characters, but whom we would not want to punish.

bit fast and loose with rules of police conduct or rules of evidence in their pursuit of the bad guy. Typically the laws they violate seem formalistic and unimportant, and the motive from which they act pure and noble. Thus the fact that the defendant commits a criminal act simply does not supply reliable evidence about the merits of his underlying character. The same is true when we try to infer motive, instead of character, from the performance of an illegal act. Not only will the law refuse to exonerate the compassionate and well-motivated mercy killer, but it will even treat him offense as a more serious one than that of the jealous lover who kills in the heat of passion.<sup>38</sup> Yet we surely do not think that a wife who helps her beloved husband end his suffering from a terminal disease necessarily acts on a worse motive than the person who kills his wife's paramour when he discovers the couple *in flagrante delicto*. We must conclude that bad motive, like bad character, is simply not a good correlate for criminal liability.

The most important difficulty with the character theory for our purposes, however, is the account it offers of action performed on the basis of a disposition or character trait. For the account of dispositions it offers makes it unsuitable for explaining defenses predicated on that notion. The problem stems from the fact that character theorists overlook a central psychological fact about character traits, namely that they produce actions semi-automatically. Once the nature of such actions is fully understood, it should be clear that it is not possible to use character in the way the character theorist proposes. The problem is that the semi-automatic nature of actions expressing a disposition makes them function like an intervening condition between the agent's formation of his character and the action that later results from that character trait. For this reason, responsibility for character does not translate into responsibility for the acts that follow from those traits.

---

38. See Model Penal Code § 210.2 & § 210.3 (Proposed Official Draft 1962).

Let us focus on the notion of a disposition, rather than on the narrower notion of a character trait. Dispositions are simply entrenched patterns of behavior; they need not be morally valenced. The salient feature of dispositions, which traits of character share, is that they are at least partly habitual, and require less involvement from the agent's rational faculties than ordinary willed action. Once a person has acquired a generous disposition, for example, the decision of whether to give to others is no longer entirely a matter of choice. Indeed it is this feature that makes it helpful to acquire dispositions in the first place. Because acts that stem from a disposition require less of an effort of will than those that are *sui generis*, having a disposition makes actions of the relevant sort easier to perform. Thus the generous person finds it easier to perform acts of generosity than the stingy person, because she has a standing set of responses in circumstances calling for generous acts. Instead of deliberating about what to do when an opportunity for generosity arises, she may often simply find herself performing such acts when the situation calls for them. And this is because she takes performing generous acts in certain circumstances as a non-optional end towards which she strives. Her practical reflections consist entirely in trying to figure out how to perform the generous act, not in *whether* to perform it.<sup>39</sup> We often capture this by saying that doing the generous thing "comes naturally" to such a person, by which we mean that the person does it with little reflection or internal debate.

It is important, however, not to obscure the fact that actions that stem from a disposition are still fully intentional, and thus ultimately remain up to the agent. An experienced driver, for example, may shift gears largely

---

39. Aristotle says something similar about the notion of character. He says, "that which is in truth an object of wish is an object of wish to the good man," and also, famously, that "We deliberate not about ends, but about means." Aristotle, *Nicomachean Ethics*, in *The Basic Works of Aristotle* 971, 970 (Richard McKeon ed., W.D. Ross trans., Random House 1941). Together, these passages suggest that the virtuous man simply takes objects of virtue as his end, without deliberating about them. His deliberation is only about how to realize the virtuous ends he takes as given.



without reflection. The absence of deliberation does not negate the driver's control over his bodily movements. His control lies in the fact that the driver remains capable of rejecting his habitual movements in favor, say, of shifting gears at an unaccustomed moment or in an unusual pattern or manner. The point of acquiring the habit is not literally to make it impossible for the driver to do anything other than perform his habitual bodily movements, but to make it substantially more difficult for him to do so. This is why the driver can rely on his ability to perform the right actions at the right time: The fact that the movements required for shifting gears have become habitual makes it easy for the driver to execute them.<sup>40</sup>

Like the experienced driver, the actions of the person with a disposition display the following combination of features: First, we are assuming that the agent is responsible for the acquisition of her own disposition.<sup>41</sup> Second, the actions to which the relevant dispositions give rise are fully voluntary and intentional. Yet, third, the actions displaying the disposition are executed largely without the agent's rational deliberation. Just as the driver would have to struggle to force himself to adopt an unaccustomed pattern of shifting, so it would take an effort of will for the generous person *not* to respond with generosity. It is for this reason, I am suggesting, that actions that proceed from a disposition are less under the rational control of the agent who performs them. While they are still fundamentally up to the agent, an individual will have to fight his own dispositions to act other than the

---

40. For a discussion of expert behavior, see Hubert L. Dreyfus & Stuart E. Dreyfus, *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (1986).

41. The assumption actually may be a questionable one. In particular, it appears to stand in some conflict with the law's assumption that the fact that a defendant had a rotten social background should entitle him to at least partial mitigation of his sentence. For if the law supposes that a defendant deserves to be exonerated for a bad act he performed on the grounds that his background shaped him in a way that made him violate the law, then the law seems to be assuming that human beings are not responsible for their characters after all. In what follows I shall assume the Aristotelian thesis throughout nonetheless, since debating this point will take us too far afield.

disposition would indicate. Alternatively, if the disposition is very well established and very controlling, the agent may actually have to alter the disposition before he can alter the course of action he would otherwise perform. Thus while agents acting under the weight of a disposition control the actions they perform, the control they exert is indirect: control over the action operates through control over the acquisition and maintenance of the disposition from which the action issues.

But if the person who acts on a disposition is still capable of acting against the weight of his disposition, why would the fact that an act stems from a disposition provide a basis for exonerating an agent who performs it? To see why dispositions might diminish an agent's culpability for a bad act he performs, consider the behavior of a drug addict trying to satisfy his pangs of addiction. On the one hand, in most cases the addict is fully responsible for his addiction, in roughly the same way that Aristotle posited for virtuous and vicious actions: The agent gave himself the disposition by repeatedly performing acts of the sort the disposition involves. The addict's bodily movements remain under his control, even when he is most fully in the grip of his addiction. Evidence for this lies in the fact that he will surely be capable of instrumental reasoning about each act he performs. He may deliberate rather carefully about where and how to administer the drug to himself, and he may even have the capacity to adopt relatively long-term plans in pursuit of his end. But despite the fact that he controlled the onset of the addiction in the first place, and the fact that he performs the actions intentionally that manifest the addiction, it is no longer up to the addict whether or not to take the drugs. There is thus a sense in which the addiction vitiates the addict's ultimate responsibility: Although he is responsible for becoming an addict in the first place, that responsibility does not transfer to particular acts performed in satisfaction of that addiction, even acts that are properly speaking intentional. For, like the generous person we considered above, taking the drug has become a non-optional end in pursuit of which he must act.

I have been arguing, against the character theorist, that even if we grant that agents are responsible for their dispositions or characters, they are not necessarily morally responsible for the acts that follow from those dispositions. This is so, despite the fact that the actions that follow from the disposition are intentional actions, and thus are at least indirectly subject to an agent's control. There is, however, an odd consequence of my argument. On the line I am suggesting, the generous person is not particularly praiseworthy for acts of generosity, and the miser is not terribly criticizable for acts of stinginess. While each may be praiseworthy or blameworthy for being the sorts of people they are, neither can be praised or blamed for *acting* as they do, *given* that their actions follow from deeply ingrained dispositions. Since the generous person cannot particularly help doing the generous thing, he is merely acting as he must in giving to others. By contrast, if the miser were to be as generous as the generous person, he would be much praised for his act, since acts of generosity would be particularly impressive coming from him, given how difficult it is for him to perform them.

I believe, nevertheless, that this is a defensible upshot of the view of dispositions I have been articulating. It is consistent with a Kantian perspective on the moral worth of human action. Recall Kant's remarks about the shopkeeper who returns the correct change to his clients.<sup>42</sup> Kant says that if the shopkeeper does this because he is a person of sympathetic temperament, his action has no moral worth. It is only when he does it against the weight of present inclination that his action has moral worth, because only then can we be sure it stems from the will and is motivated by the thought of duty. Kant thinks that only actions performed against the weight of an agent's inclinations can have moral worth, since otherwise an agent's dutiful acts are motivated by pleasure. Such acts

---

42. Immanuel Kant, *Groundwork of the Metaphysics of Morals* 65 (H. J. Paton trans., Harper & Row 1964).

would not be entirely freely chosen; they would, in Kant's terms, be "heteronomous," rather than fully "autonomous."

In a similar vein, I am arguing that actions performed against the weight of an agent's dispositions are most directly the product of an agent's deliberative faculties, and so are the actions for which he is most particularly responsible. While character theorists are intent on seeing action that stems from standing dispositions as *most* emblematic of moral agency, I am suggesting that we have reason to think of dispositions as standing in opposition to rational agency, insofar as the action they produce requires less from the will of the agent than would an action chosen for its own sake. If this is correct, then it is action that is *out of character* for the agent that most bears the stamp of its author's moral agency, because such action is most chosen. The result is problematic for the character theorist's approach to the excuses. If the character theorist were correct that we could infer nothing about the character of the person who participates in a crime at knife-point, that would tend to inculcate him for the crime, rather than suggest a reason why he should not be held responsible. Moreover, if we could draw an inference from bad act to bad character trait, that would tend to exculpate, rather than to inculcate. So the character theorist appears to have matters exactly backwards when he claims that the reason we give an agent an "excuse" is that the action he performs is not emblematic of his character. He is, if anything, *more* responsible for action that is out of character than for action that is in keeping with his deepest traits.

Having seen that the character theory will not provide us with a satisfactory approach to criminal responsibility, let us proceed to consider the nature of the rational excuses more specifically. As we shall see, we must deploy dispositions in quite a different way from the way the character theorist we have just considered does.

## V. RATIONAL EXCUSES AND ADAPTIVE DISPOSITIONS

The account of the rational excuses I shall develop is one I presented in preliminary form in two earlier articles, one dealing with the duress defense and the other with self-defense.<sup>43</sup> In those articles, I argued for a reversal of the usual appeal to character proposed by the character theorist: Duress and self-defense, I claimed, do not exonerate because people who act under these kinds of pressures behave in ways that are "out of character." Instead, it is precisely because such acts stem from settled dispositions in the agent that we are inclined to allow a defense under these circumstances. I argued that we exonerate in such cases because the act the agent performs in violation of the law stems from standing features of his motivational set, features we are prepared to endorse. Duress and self-defense are cases in which we cannot eliminate the legal violation without eliminating the disposition from which it arose. In this Part, I shall attempt to generalize the arguments I made in the context of these specific defenses to the wider class of rational excuses.

On the theory of the excuses I would like to propose, rational excuses are cases in which the question of the agent's responsibility for the criminal act refers us to the larger choice of disposition from which the agent acted. We can explain defenses like duress, self-defense, entrapment, and personal necessity along these lines, because such defenses constitute claims that the disposition that produced the defendant's wrongdoing is itself socially beneficial. Thus the reason we afford defendants an excuse in such cases is that we do not want to discourage individuals from adopting the relevant dispositions. We shall see presently why this is the case.

The account I am presenting is largely consistent with the choice theorist's account. For it accepts the choice theorist's basic premise that human beings are responsible

---

43. Finkelstein, *supra* note 10.

for conduct in which they choose to engage. But it recognizes, first, that a certain category of defenses calls into play the notion of a disposition, and second, that choice operates differently where a disposition is at issue than where it is not. The person who acts on the basis of a disposition exercises choice, but usually not with respect to individual actions. Instead, he must choose between opposing or reinforcing the disposition, that is, he either stands in the way of the disposition or continues to allow it full force. In a case in which a person acts on a bad disposition, his responsibility for the ensuing bad act is based on a kind of omission—similar to the choice a person might make not to oppose another agent who was about to do something terrible. Of course a person is not responsible for all the results of his inaction. He is only responsible for harmful results that occur through his inaction if he had a duty to prevent those results. A person has a duty to prevent the results of his own bad dispositions where the operation of that disposition will result in a bad act. Assuming, then, that it is possible, however difficult, to refrain from acting on a disposition, a person is responsible for the bad acts he performs that result from bad dispositions, because he has a duty to oppose even a semi-automatic process of which he is the cause and which he controls. But he has no obligation to stop himself from acting on a *good* disposition, even when it produces a bad act, because it is morally commendable that he has that disposition. In particular, the law grants a defense in such cases, in recognition of the fact that if the agent were to stop himself from acting on the disposition every time it produced a bad act, he would be in danger of losing the disposition altogether.

The dispositional account of the rational excuses will diverge from the choice theorist's account in relatively few places, since the instances in which dispositions play a role remain relatively few. Disposition is still irrelevant to the theory of inculcation, and it is irrelevant to defenses other than the rational excuses. Thus the theory does not extend to excuses involving involuntary conditions, such as

epilepsy, or to conditions that impair an agent's rational faculties, such as infancy, insanity, or involuntary intoxication. The choice theorist is correct to think of these as lack of capacity defenses, since a person who has a tendency to suffer epileptic seizures cannot be said to have an "epileptic disposition;" nor can a non-rational child be thought of as having a disposition to act irrationally or non-rationally. Conditions that impair capacity are preclusive of the exercise of dispositions, since dispositions operate through, rather than in lieu of, an agent's rational faculties. The theory also does not affect the analysis of what we might call "true justifications," namely cases in which the agent claims a privilege based on the impersonal benefits he supplies for society at large. For these are cases in which the defendant's act, considered in and of itself, is commendable. His defense therefore does not depend on the merits of the disposition from which he acts.

Thus far my reference to "good" dispositions has been entirely generic. We must now consider in what sense the dispositions the rational excuses reflect are "good," as well as to address the question of why the law might want to privilege them. My more specific claim is that the rational excuses involve what I shall call "adaptive" dispositions, meaning that they are dispositions an agent acquires in pursuing his own welfare, but which generate collective gains for members of society as a whole. That is, to the extent that everyone adopts the dispositions in question, there will be collective gains that leave each person better off than he would be in the absence of those dispositions. My argument is that because there are gains to everyone from the widespread adoption of these dispositions, we often tolerate their operation when they lead individuals to violate the law.

To illustrate, consider the fact that people who acquire dispositions to keep promises will be able to enter into cooperative ventures with one another for mutual benefit. The well-known example of the Humean farmers will make this clear. Alfred's field is ready for plowing this week, and Bertram's field will be ready next week. Neither farmer

can plow his field by himself. Alfred proposes that Bertram help him plow his field this week, and in exchange Alfred will help Bertram next week. If Alfred can convince Bertram that he intends to keep his promise to plow Bertram's field next week, it would be rational for Bertram to agree to Alfred's proposal, since both will be better off than if they do not enter into such an agreement. But can Alfred convince Bertram? In answering this question, we should assume that there are no further benefits to Alfred from plowing Bertram's field. That is, Alfred would gain no reputational advantage by doing so, nor will there be any future course of dealings between Alfred and Bertram that would make it advantageous to either to cooperate now. In addition, we assume there is no legal or other coercive enforcement of any agreement the two farmers might make.

In the absence of reputational incentives or other external sources of benefit from cooperation, it is hard to see how it could be to Alfred's benefit to make good on his promise to help Bertram next week. For once Bertram has already helped Alfred, there will be no further benefit to Alfred from keeping his promise. But if this is so, and if Bertram is rational, and he and Alfred each has knowledge of the other's rationality, Bertram will not help Alfred. The result is that both will be left worse off than if they were able to cooperate. Suppose, on the other hand, that Bertram knows that Alfred is a very reliable promise keeper, and that he keeps promises even when it is not to his benefit to do so. Then Bertram will be more inclined to believe Alfred's promise, and so will cooperate with him. In cases of this sort, then, there will be mutual gains from the acquisition of a disposition. If agents can acquire dispositions like promise keeping, they will be able to make use of them to convince one another to enter into otherwise unenforceable cooperative arrangements that accrue to the benefit of both. It is for this reason that dispositions to keep promises and to cooperate with others are "adaptive," despite the fact that their operation does not always



maximize an agent's welfare in the context of the choice the agent must make.<sup>44</sup>

Notice that adaptive dispositions function something like pre-commitment devices. A pre-commitment device is an external mechanism for ensuring that people keep their commitments, which works by altering the payoffs to remove any benefits from violating those commitments. A disposition accomplishes the same result by way of *internal* commitment instead. Unlike an actual pre-commitment device, however, an adaptive disposition achieves compliance without exacting a penalty for non-compliance. An agent who acquires an adaptive disposition has "internalized" the penalty for non-compliance that external pre-commitment devices supply. She feels moved to honor her commitments without the fear of the kind of sanction an external pre-commitment device would impose. The disposition works to produce compliance with a social norm because it acts as a filter on the admissible options from which an agent who possess the relevant disposition can choose.<sup>45</sup> Moreover, a disposition is in some respects preferable to a pre-commitment device. In particular, pre-commitment devices have costs that dispositions do not appear to have. And if individuals can achieve cooperative behavior in the absence of incurring such costs, there will be a surplus for the parties to divide that potentially leaves both parties better off.

To see the costs of pre-commitment devices more clearly, suppose in order for Alfred to convince Bertram he intends to make good on his promise to help him plow his field, Alfred must hire someone to break his kneecaps if he fails to render the promised assistance. He must pay the enforcer \$50, let us say. At that price, it is still worth it to Alfred to enter into the agreement with Bertram, since the

---

44. For an argument that reciprocating in such cases in the absence of pre-commitment strategies, reputational effects, or other exogenous pay-offs from cooperation, see David Gauthier, *Resolute Choice and Rational Deliberation: A Critique and a Defense*, 31 *Noûs* 1 (1997); David Gauthier, *Assure and Threaten*, 104 *Ethics* 690 (1994).

45. Michael E. Bratman, *Intention, Plans and Practical Reason* (1987).

benefit he will receive from having help with his own field will more than make up for the cost of the pre-commitment. But clearly Alfred would be better off if he could convince Bertram of his sincerity without having to pay the enforcer. If he had a promise-keeping disposition, he would reap the benefits of cooperation with Bertram, without having to pay enforcement costs. And if acquiring the relevant adaptive disposition is truly something agents control, then it would be irrational for someone to proceed by way of pre-commitment instead. Thus I am suggesting that we tolerate violations of our primary norms to make room for the operation of dispositions because if we did not, it would be much more costly for us to achieve cooperative behavior. For we would have to make use of legal norms instead of informal mechanisms, and such mechanisms would involve substantial additional enforcement costs.

In the context of the rational excuses, I am suggesting that dispositions like loyalty to loved ones or strongly self-preferring commitments are adaptive, in the sense that there are collective gains from the possession of those dispositions overall, despite the fact that they sometimes lead agents to break the law. It might be better if we did not allow a defense to individuals who are threatened with death or serious bodily injury, since at the margins that will result in more coercion, and hence in more crime. And in many cases, it will result specifically in a greater evil rather than a lesser one, since a person may have a duress defense even if the danger with which he is threatened is not graver than the criminal act he is asked to perform.<sup>46</sup> But we could not penalize the legal violation in this case without also penalizing the possession of the disposition

---

46. While the defense of duress is generally not allowed as a defense to murder, many commentators think the restriction makes little sense. See, e.g., Wayne R. LaFare and Austin W. Scott, *Criminal Law* 438 (2d. ed. 1986) ("It is not proper, on principle, to limit the defense of duress to situations where the instrument of coercion is a threat of death or serious bodily injury. A threat to do bodily harm less than serious bodily harm, or a threat to destroy property or reputation, ought to do . . ."); see also 2 Robinson, *supra* note 16, § 177(c)(2), at 354 (arguing that the relevant consideration is "the nature of the actor's state of coercion and his ability to resist it").

that led to that violation. And since there are cooperative gains from the underlying adaptive disposition, we do not wish to discourage the disposition in this way.

But someone might offer the following objection to the argument I have been making. The fact that there are mutual gains from allowing agents to develop adaptive dispositions does not seem to require us to accept the operation of such dispositions in instances in which the disposition is *not* socially advantageous. That is, if we would truly prefer that a person who possessed adaptive dispositions followed the primary legal norm, why not use legal rules to police the line between the instances of, say, loyalty we want and the instances of loyalty we do *not* want? Why must we suffer with the results of loyalty in a case in which that sentiment leads an agent to violate one of the law's prohibitory norms, such as the norm against killing?

The answer is that the operation of the disposition cannot be policed so precisely. A disposition is an imprecise mechanism for bringing about socially cooperative behavior. On the one hand, the costs of trying to enforce the disposition in non-legal contexts would be much too high. Consider the "costs" of attempting legally to enforce a disposition like loyalty in every instance in which we would like one another to be loyal. Social commitments, marriage promises, gratuitous exchanges among close friends or family would all become the subject of legally enforceable exchanges. Forming dispositions is a preferable way of binding one another than explicit pre-commitment devices, since the latter involve costs that achieving cooperative behavior by way of dispositions would not. On the other hand, it is also very difficult to suppress the disposition in those instances in which we would prefer it did not operate. Thus in order to be able to rely on the correct operation of the disposition in those instances in which we want them, we must tolerate their operation in some cases in which we do not. Where an adaptive disposition conflicts with a legal prohibitory norm, we sometimes exempt the agent for his responsibility for having violated the latter.

In particular, the adaptive dispositions that are of relevance to explaining the rational excuses are among the most fundamental to mutually beneficial social arrangements. They are dispositions like self-preservation, which lead a person to prefer her own life over the lives of others. Hence we allow a defense of self-defense, even in cases in which social welfare would be better served if the defendant could forego the conduct by which he violates the law, because, for example, more innocent blood would be shed if the agent acted on the self-preserving disposition than would be if the agent acted otherwise. In cases like that of the child attackers, society might be better served if the defendant chose not to save his life. But a society of mutual benefit cannot function if individuals do not seek to preserve themselves and enhance their own well-being. Personal necessity, such as the case of the men clinging to the plank, is also a clear example of agents violating a prohibitory norm because they are disposed to choose their own survival over the lives of others. Once again, it is a disposition we endorse, even though society would presumably have no objection, and indeed might prefer, if a person chose to sacrifice himself to save another under the circumstances. Indeed, the disposition of prudence is so fundamental to a society founded on principles of mutual advantage that self-preserving acts may seem to be closer to justifications than excuses. That would explain why self-defense and necessity have traditionally been considered justifications rather than excuses, despite the fact that social welfare is not always directly enhanced when agents engage in self-preferring acts.

The exception to the overwhelming importance of self-preferring dispositions is when agents are motivated to protect members of their family or others to whom they have developed special attachments. Thus the defenses of duress and provocation are sometimes issued to make room for other-oriented dispositions, even when they go awry and lead individuals to kill those they love. Not every case of duress or provocation takes this form. In many instances, the defendant is acting on a self-preferring

motive, as is the case when the coercion he experiences threatens him with death or bodily injury if he does not commit a crime, or when a person is provoked by insults that incite his self-protective feelings. But where the coercion operates by threatening the well-being of a loved one, or where a person is provoked by feelings of jealousy, the dispositions in question are other-oriented.

Entrapment is similarly a defense in which the defendant's willingness to violate the law is arguably other-oriented. But here the disposition operates in an attenuated sense. A defendant's susceptibility to the blandishments of a police officer or other law-enforcement agent is an understandable result of his being the sort of person who cooperates readily with others. He is subject to persuasion because he is a social being, and he wants to be helpful and cooperative with a person who invites his assistance or encouragement. Admittedly, there are complications here. It is puzzling, for example, that the defense of entrapment is limited to cases in which the defendant receives pressure from a governmental agent. For if the entrapment defense were understood as accommodating a social disposition in the way I have described, one might expect similar pressure from a non-governmental official to entitle a defendant to a defense as well. It is at least in part for this reason that some commentators are inclined to see entrapment as a non-exculpatory defense, similar to the diplomatic immunity or double jeopardy defenses.<sup>47</sup> And it must be admitted that some applications of the defense seem to fit this model best, such as cases where the defendant is afforded a defense because the government has engaged in conduct that is simply too reprehensible to allow it to prosecute the defendant for *his* bad conduct.<sup>48</sup> Probably the defense should be thought of as non-exculpatory in these kinds of cases. But the heart of the doctrine seems taken up by cases in which the defense *is* exculpatory: the reason the

---

47. See 2 Robinson, *supra* note 16.

48. I am grateful to Peter Westen for emphasizing the distinction between the different kinds of entrapment cases.

defendant merits exoneration in such cases is that the defendant would not have violated the law without pressure from the governmental agent. In such cases, it is plausible to explain the defense by saying that the defendant is manifesting a cooperative disposition. And since the disposition is in evidence in cases in which the instigating agent is a private citizen, we should probably extend the defense to such cases.<sup>49</sup>

A point of clarification before I close this section. It is common to think of excuse as applying to actions society would prefer were not performed, actions where we would ideally wish the agent to abide by the law, but where we exonerate him for violating it. By arguing that some cases of self-defense, some cases of necessity, duress, entrapment, and other defenses should be conceived of as "rational excuses," I may appear to be suggesting that we would *prefer* the individual attacked by small children to stay his hand and not to kill them in defense, or for the starving individual to refrain from stealing the loaf of bread. I have indeed written thus far as if this were the case, by saying that the law "tolerates" a violation of its primary norms in cases in which the violation stems from an adaptive disposition. But I must now qualify that claim somewhat. In some cases it is clearly correct that we are prepared to grant a defense or partial defense, despite the fact that we would "prefer" that the defendant have conformed to the prohibitory norm. This would most obviously be the case in instances where the provocation defense might apply. But it is a much more tenuous description in self-defense, personal necessity, or duress cases, where we normally do not think the defendant ought to sacrifice himself to save the lives of innocent attackers, to avoid violating a prohibitory norm that does not pertain to loss of life, or to avoid committing a crime. We do not necessarily wish that the defendant allow himself or his

---

49. In this regard, however, the dispositional approach I am proposing here does not lag behind character theory approaches, since the character theorist cannot distinguish good character from bad according to the identity of the entrapping agent.

loved ones to die in order to avoid assisting in the commission of a crime. Indeed, we often recognize the defendant's entitlement to defend his own life or the lives of others. This is most obviously the case, for example, in self-defense cases, where we tend to think the defendant is acting from a claim of *right* in defending his life, even if he must threaten the lives of innocent attackers in order to do so. How, then, is this sense of entitlement compatible with the argument from rational excuse I have offered?

To say that a defendant should be exonerated on the basis of a "rational excuse" is not necessarily to say that he has done something society regards as wrong or reprehensible, although some instances of rational excuse fall in this category. It may be that from a moral standpoint we are uncertain or ambivalent about the moral merits of the defendant's conduct. It is true that the category of rational excuse does not apply to conduct that the law wishes to exonerate because it is particularly commendable. Acts of this sort, such as true lesser evils cases, belong to the category of justification proper. But the fact that the conduct is forbidden by law and that it is not particularly commendable does not mean that it is reprehensible either. Thus the category of "rational excuse" is not restricted to morally opprobrious conduct. It should also be stressed, however, that there is a common misconception that may be driving the view of excuse as applying to morally reprehensible action, and this is the idea that action we have a moral right to perform is *commendable*. Thus in support of seeing self-defense as a justification, it is sometimes said that surely people have a right to defend themselves, even against innocent attackers, and then to add that it is a good thing when people stand on such a right. While people may have a moral right to defend their lives, it hardly follows that it is always a good thing that they exercise this right. Some rights are morally indifferent, and some can be positively loathsome when exercised. Thus the fact that the legal right of self-defense may be supported by a moral right of self-defense does not seem to me to count in favor of

treating the legal right as a justification. And, on the other hand, to treat conduct in self-defense as excused does not tend to suggest that it is bad or morally reprehensible.

I have argued that the defenses I have called "rational excuses" are best understood as instances where the law exonerates because the defendant's choice to violate the law reflects a disposition the law wishes to endorse. Strong prudential or other-regarding motivations sometimes lead an agent to violate the primary norms of conduct, and where this is the case, we do not hold the agent to the same standards we would absent such motivation. We must now consider the justification for constructing our rules of legal defense according to the dictates of rational self-interest, and so instituting rational excuses as exceptions to basic legal norms. That is, if the primary legal norms establish a standard of conduct below which we do not want individuals to fall, then why should we exempt individuals from those rules to make room for adaptive dispositions, absent a justification for the exempt conduct? I shall address this and other issues in the concluding section below.

#### VI. CONCLUSION: A CONTRACTARIAN JUSTIFICATION FOR THE RATIONAL EXCUSES

If the possession of the foregoing dispositions is of benefit to each individual, we have a basis for thinking of the exoneration in cases in which a rational excuse applies as justifiable: Given that it is not possible to ensure that people act cooperatively in the absence of strong cooperative dispositions, there are reasons of mutual advantage for excusing an agent who acts on the basis of an adaptive disposition in cases in which doing so leads her to violate the law. Rational agents may not have reason collectively to prefer the particular illegal action that issues from an adaptive disposition in a given case, but they have reason to prefer that others possess such dispositions. Rational agents inventing rules governing the operation of the excuses would choose to exonerate others who act on



such dispositions, since such rules would leave each person individually better off than he would be in the absence of such rules. And if this is the case, then rules of this sort would be readily agreed to by members of society selecting the legal rules by which they wish to be governed. The rules governing rational excuses are thus fair and justifiable, because they represent the choices rational agents attempting to enhance their own well-being would make for themselves in an *ex ante* position of agreement.

Consider once again the example of duress. As I suggested above, it may well be better for society as a whole if legal rules insisted that innocent individuals not participate in the criminal activities of those by whom they are coerced. This is probably even true where requiring an innocent agent to resist the demand to participate in criminal activity would have serious personal consequences. One reason to prefer sticking to the primary legal rule over the personal exemption is obvious: potential coercers would have less chance of successfully coercing others if the potential victims knew there would be no criminal defense for their behavior. Thus adherence to the legal rule may appear to be optimal here. Yet I have argued that rational individuals would agree to exempt victims of coercion from their responsibility for the criminal act in this sort of case. Why? I have argued that the non-legal benefits of adaptive dispositions give agents reason to prefer a society in which such dispositions are given wide deference to one in which they are severely restricted. Thus even if there are good reasons for sticking with the primary legal norms, rational agents may regard themselves as better off under a regime that carves out various exceptions to those norms. This may even be true if sticking to the norms would actually be better, from the standpoint of social utility, to allowing the exceptions, even taking the benefits of the dispositions into account. For the fact that a rule would maximize social welfare does not mean that each individual who must live under that rule

has reason to think his own welfare maximized by accepting its terms.<sup>50</sup>

There is, however, a final wrinkle we should briefly explore. Earlier we made the argument that dispositions are preferable to legal enforcement, because there is a cost-savings from avoiding the costs of pre-commitment. But arguably dispositions involve costs of their own. And if so, then the argument we saw about the superiority of dispositions over pre-commitment devices seems to apply to non-dispositional over dispositional decisionmaking: While a disposition allows agents to reap the benefits of cooperation with other agents, it would be better if human beings simply committed themselves to behaving cooperatively. So if dispositions are preferable to pre-commitment devices because they accomplish the same ends without the costs, then surely rational, case-by-case commitment is preferable to dispositional commitment, for the same reason. If so, then the contractarian argument we have just considered for indulging dispositions is subject to doubt.

It is admittedly true that dispositions involve costs. There are costs involved in acquiring the disposition in the first place. And there are costs of the sort associated with any practical rule, stemming from the fact that rules are imperfectly tailored to the situations to which they apply. Thus the generous person might continue to give even after his recipient has shown himself to be an undeserving wretch; the miser might continue to pinch pennies even after the need to do so has been eliminated by his winning the lottery. And the person who is deeply loyal to family

---

50. This last point is an exceptionally complicated one, and I have mostly side-stepped it in the above argument. That is, I have assumed for the most part that making room for adaptive dispositions as exceptions to primary legal norms would maximize society's overall welfare, as well as maximize the welfare of individual agents. But at least theoretically, it is possible that each agent in a position of antecedent social choice would regard himself as better off under a legal regime that was non-optimal from the standpoint of social welfare. And in that case, the contractarian justification for legal rules I am suggesting here would diverge from the utilitarian or welfarist analysis that would support sticking with the legal rule.

and friends may commit a crime as a result of such loyalty, under circumstances in which it would be better if he did not. Like any practical rule, adaptive dispositions will sometimes be both over and under-inclusive.<sup>51</sup>

Thus a society of *perfectly* rational agents seeking the benefits of cooperation and coordination would not make use of adaptive dispositions. Individuals in such a society would simply cooperate where there were mutual gains from doing so, and otherwise they would proceed by enforceable legal prohibitions where those were beneficial instead. That is, perfectly rational agents would ask themselves whether they would be better off under a given cooperative arrangement than they would be outside that arrangement. Where the agreement provides a net benefit, perfectly rational agents would be faithful to their agreements: they would use internal commitment to reproduce the effects that less than perfectly rational agents would achieve through the formation of dispositions. And where this was the case, the contractarian argument for the incorporation of the rational excuses into legal rules would not go through. For agents would not form dispositions under these circumstances, and legal rules would not have to be shaped to accommodate them. But the human actors for whom the rules of the criminal law are fashioned are not perfectly rational. They are not agents who would otherwise be capable of instituting cooperative ventures in the absence of enforcement or some reputational or other incentive to cooperate. Cooperative dispositions, on the other hand, make it possible for agents to cooperate in the absence of such external incentives. Thus the contractarian argument justifying the existence of the rational excuses we have just considered is compelling as long as we understand its applicability as limited to legal societies established by somewhat imperfectly rational agents.

---

51. See Frederick Shauer, *Playing By the Rules: A Philosophical Examination of Rule-based Decision-making in Law and Life* (1991)

The significance of the account of the adaptive dispositions I am proposing should now be clear: Dispositions are strategies agents can adopt in order to compensate for the fact that they are less than perfectly rational. They allow imperfectly rational agents to approximate the position they would be in if they were perfectly rational. The law of criminal defenses takes that imperfection into account, allowing basically rational agents to compensate for their own irrationality by forming dispositions. Such dispositions are "adaptive," because they allow human beings to act in ways that ultimately increase their own well-being and those of their fellows. The criminal law accommodates and encourages the formation of such dispositions by exonerating or partially excusing agents who act on the basis of them.