THE CONTINUITY OF JUSTIFICATION DEFENSES

Kyron Huigens[*]

This Article presents a legal theory of legal justification. It examines the elements of offense definitions and justification defenses—or, more accurately, the conjunctions and disjunctions between these elements—to resolve a longstanding issue in the theory of legal punishment. The unjustified actor who believes she is justified seems to deserve an acquittal, while the justified actor who does not know he is justified seems to deserve conviction. But we face a dilemma: we seem to have to acquit both or convict both. This Article shows a principled way to rule the mistakenly unjustified actor within, and the mistakenly justified actor out of, the class of justified actors, thus matching our legal judgments to our moral judgments concerning these cases.

The route to this conclusion is as important as the conclusion itself. Much of the argument employs tables of elements and the conjunctions and disjunctions between them. These tables are illuminating—surprisingly so for such a simple device. This elements analysis serves to demonstrate some elementary points, such as the existence of a genuine difference between offenses and justification defenses. And it can demonstrate more complex and surprising points, such as the unavailability—on any view of wrongdoing and justification—of attempt liability as a second best solution to the dilemma presented by mistakes in justification.

The Article also uses Aristotelian punishment theory as a complement to elements analysis. The use of virtue ethics is not simply a matter of equating solutions to moral issues with solutions to legal issues. That would not be the legal analysis of a legal problem that is promised here. Instead, the justification defenses are portrayed as the product of the specification of competing ends—a distinctively Aristotelian model of deliberations on ends—making the case that the justification defenses, even taken as permissions that cancel the prohibi-

* Professor of Law, Benjamin N. Cardozo School of Law, Yeshiva University. I wish to thank Tony Dillof for helping me to clarify my thoughts concerning the Model Penal Code's conception of justifications. I also wish to thank Matthew Shroyer, Cardozo School of Law class of 2010, for his research assistance.

tion for reasons that conflict with the prohibition, are continuous with criminal offenses.

I. INTRODUCTION

In the theory and practice of legal punishment, the element is a notional tool on the order of the wheel or the lever. The codification of criminal law serves fundamental rule-of-law values, and the elements of offenses and defenses are the essential constituents of a criminal code. The project of codification inevitably brings deep moral and theoretical issues to the surface, and many, if not most, of these issues arise in connection with the definition, choice, and arrangement of offense and defense elements. Little of the theory of legal punishment, however, makes deliberate use of the element as an organizing principle of inquiry. As a result, questions in legal theory often are framed as questions in moral theory. This is a mistake, and this Article will avoid it. I will use an elements analysis to shed some light on the structural relationship between offenses and justification defenses. This elements analysis will be complemented by some concepts and a method of analysis found in an Aristotelian, or aretaic, theory of punishment.[1] The combination of these two approaches will solve a basic problem in the theory of legal punishment.

The issue is whether an actor who makes a mistake about justifying circumstances is genuinely justified. There are two distinct cases: the mistakenly justified actor who believes that he is not justified, and the mistakenly unjustified actor who believes that she is justified.[2] These cases give rise to a dilemma. As a moral matter, the mistakenly justified actor deserves punishment and the mistakenly unjustified actor does not. As a legal matter, it seems that we must either acquit both or convict both. Most of the literature on this problem approaches it from a moral point of view, attempting to account for our different judgments of the two cases. I find this odd; the dilemma is a legal dilemma, not a moral one. Our moral judgments of the cases are relatively clear. The problem is that they baffle the law. The recommended legal solution to the problem is to permit the cases to be decided on the actor's belief alone, regardless of the facts. But this solution departs so far from both our ordinary understanding of what it means to be justified and our rule-of-law values, that it only offers us a trilemma to take the place of our dilemma.

1. *See generally* R.A. Duff, *Virtue, Vice, and Criminal Liability: Do We Want an Aristotelian Criminal Law?*, 6 BUFF. CRIM. L. REV. 147 (2003); Kyron Huigens, *On Aristotelian Criminal Law: A Reply to Duff*, 18 NOTRE DAME J.L. ETHICS & PUB. POL'Y 465 (2004) [hereinafter Huigens, *Aristotelian Criminal Law*]; Kyron Huigens, *Virtue and Inculpation*, 108 HARV. L. REV. 1423, 1480 (1995) [hereinafter Huigens, *Virtue and Inculpation*].

2. These are terms of art in which "justified" and "unjustified" refer only to the factual configuration of the case and its outcome under the factual elements of self-defense—and not to its conception, analysis, or ultimate outcome. Whether an actor ultimately is genuinely justified, guilty as charged, excused, guilty of an attempt, and so on, depends on further considerations—primarily the actor's criminal fault or culpability—that define the terms of the theoretical debate.

Sorting out the moral basis of our judgments might eventually serve to resolve the legal trilemma.[3] This Article will take the more direct route of addressing the legal trilemma with an analysis that stresses the formal features of offenses and justification defenses.

The "continuity" in my title is continuity in legal desert for criminal punishment—more specifically, continuity in the proof of criminal offenses and justification defenses. This is a formal and conceptual continuity having to do with the proof of criminal fault. It is not a moral continuity that might, say, function as a common ground for our judgments on the two kinds of mistake. The continuity of the justification defenses is a legal structure. Significantly, it is a positive legal structure—not a deep grammatical structure that might premise an ontology of legal wrongdoing.[4] Because it is a matter of positive law, the structure of proof of fault in offenses and justification defenses can be altered. Because of this, we have some hope of resolving the legal dilemma that is embedded in this structure. However, the continuity of the justification defenses is the product of strong formal, conceptual, doctrinal, and historical currents in the law. Charting these currents is difficult; redirecting them is unlikely. These are, nevertheless, the objectives of this Article.

Respected scholars have argued that the mistakenly justified actor is not genuinely justified,[5] that he is genuinely justified,[6] and that he is genuinely justified but guilty of an attempt.[7] I will argue that he is genuinely justified and cannot be guilty of an attempt—which is not, of course, to say that he is morally justified. As for the mistakenly unjustified actor, the consensus view is that he is excused.[8] I confess that I have no idea what this means.

3. *See* Anthony M. Dillof, *Unraveling Unknowing Justification*, 77 NOTRE DAME L. REV. 1547, 1553 (2002) ("An examination of the moral concept of justification would elaborate under what conditions a person who prima facie deserves punishment should not be punished because she is 'justified.'").

4. *Cf.* Stuart P. Green, *The Universal Grammar of Criminal Law*, 98 MICH. L. REV. 2104, 2113 (2000) (reviewing GEORGE P. FLETCHER, BASIC CONCEPTS OF CRIMINAL LAW (1998)) ("Rather, what Fletcher seems to be claiming is simply that there is, in his words, some 'underlying unity among diverse systems of criminal justice,' some 'basic design' that is common to all of these various and disparate systems.").

5. *See, e.g.*, GEORGE P. FLETCHER, RETHINKING CRIMINAL LAW § 7.4 (1978); J.C. SMITH, JUSTIFICATION AND EXCUSE IN THE CRIMINAL LAW 41–44 (1989); Russell L. Christopher, *Unknowing Justification and the Logical Necessity of the* Dadson *Principle in Self-Defense*, 15 OXFORD J. LEGAL STUD. 229, 251 (1995).

6. MICHAEL S. MOORE, ACT AND CRIME: THE PHILOSOPHY OF ACTION AND ITS IMPLICATIONS FOR CRIMINAL LAW 180–81 (1993) (the unknowingly justified actor "has done no wrong"); Paul H. Robinson, *Competing Theories of Justification: Deeds Versus Reasons*, *in* HARM AND CULPABILITY 45, 47 (A.P. Simester & A.T.H. Smith eds., 1996) ("[W]here a person's conduct in fact avoids a greater societal harm but the person is unaware of this, the conduct is justified despite the actor's ignorance.").

7. *See, e.g.*, Heidi M. Hurd, *Justification and Excuse, Wrongdoing and Culpability*, 74 NOTRE DAME L. REV. 1551, 1566–67 (1999); Paul H. Robinson, *A Theory of Justification: Societal Harm as a Prerequisite to Criminal Liability*, 23 UCLA L. REV. 266, 289–91 (1975).

8. The mistakenly unjustified actor is taken to be unjustified by definition. *But see* Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1908–09 (1984) (the mistakenly unjustified actor is "warranted"). Given that she is unjustified but apparently

John Gardner says that excuse is a "rag-bag" of ideas.[9] He takes one of these ideas—non-responsible agency—out of the bag because it does not belong there.[10] It certainly does not apply to one who makes a mistake about justifying circumstances. Gardner then fails to notice that the rag-bag is empty. The idea that he thinks is in there—success in meeting normative expectations—escaped a long time ago (in the company of the mistake of fact defense) and now resides more comfortably elsewhere, in the definition of offenses and defenses.[11]

The parable of the rag-bag suggests that the widespread and wholly incoherent idea that a mistake regarding a justification results in an excuse should be abandoned in favor of the view that such a mistake results in genuine justification. Success in meeting normative expectations, the gist of excuses as Gardner describes it, is actually the gist of criminal fault, or non-fault, to point it in the same direction as "excuse."[12] If one who makes a mistake about justifying circumstances is "excused" in any meaningful sense, it is because she is not at fault. To see these inquiries as pertaining to criminal fault instead of an excuse implies that a mistake of fact regarding justifying circumstances results in a non-crime, just as a mistake of fact regarding offense elements does. And to say that a mistake about a justification results in a non-crime is to say that the actor is genuinely justified.

We know that a mistake about justifying circumstances results in a non-crime because any similar mistake that bears on the elements of an offense results in a non-crime.[13] The Model Penal Code conceptualizes justification defenses in this way,[14] and in doing so it captures an important aspect of the common law of crime and early codified law on justification defenses. Malice, the mother of all fault criteria,[15] operates conti-

innocent, the assumption seems to be that she therefore must be excused. *See, e.g.*, FLETCHER, *supra* note 5, at 564–65; George Fletcher, *The Right Deed for the Wrong Reason: A Reply to Mr. Robinson*, 23 UCLA L. REV. 293, 309–10 (1975); *see also* Robinson, *supra* note 6, at 283–84 (the mistakenly unjustified actor is excused); Hurd, *supra* note 7, at 1565 (same).

9. John Gardner, *In Defense of Defenses*, *in* FLORES JURIS ET LEGUM: FESTSKRIFT TILL NILS JAREBORG 251, 257 (2002) [hereinafter, Gardner, *Defenses*].

10. *Id*. at 258.

11. *See infra* Part II.B.

12. I use the term "fault" in place of the more common terms "mens rea" and "culpability." The former term, "mens rea," suggests that criminal fault is limited to mental states, which is obviously untrue. The latter term, "culpability," is doubly ambiguous. It is commonly used to describe criminal liability and responsible agency, in addition to fault in offenses.

13. MODEL PENAL CODE § 2.04 cmt. 1 (1962) ("To put the matter as this subsection does is not to say anything that would not otherwise be true, even if no provision on the subject were made . . . ."); PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 62(b) (1984) ("If the defendant's ignorance or mistake makes proof of a required element impossible, the prosecution will necessarily fail in its proof of the offense."); GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART § 65 (2d ed. 1963) (the negation of fault elements by beliefs to the contrary "is not a new rule; and the law could be stated equally well without reference to mistake").

14. *See infra* Part III.C.

15. This means that malice is an ancient doctrine, see Rollin M. Perkins, *A Re-Examination of Malice Aforethought*, 43 YALE L.J. 537, 544 (1934) ("Moreover, the patent roles of Henry III point to this as a common formula in pardons granted in the 1200's to those who had committed homicide by

nuously through offense and justification defense.[16] When malice is specified into discrete, narrowly defined fault elements, these elements also operate continuously through offense and justification defense. This continuity view of criminal wrongdoing entails the genuine justification thesis. If we chart out the fault and fact elements of offenses and justification defenses, paying close attention to the conjunctions and disjunctions between them, we can see this continuity—but only if the chart is correctly interpreted.[17] The difficulty is that the chart seems initially to point in the other direction, toward the discontinuity of offense and justification defense. And this view dovetails with conceptual reasons to think that offenses and justification defenses are discontinuous. Foremost among these reasons is the fact that a justification defense is a legal permission that cancels the operation of the offense, and does so for reasons that conflict with the reasons that define the offense. An aretaic analysis of the kind I mentioned above complements the elements analysis, provides an interpretation of the chart that accommodates this conceptual objection, and confirms the continuity view of criminal wrongdoing.[18]

The argument for the continuity view does not go through easily for reasons beyond these formal and conceptual difficulties. The implicit genuine justification thesis has counterintuitive implications, and this, for most scholars, drives mistakes in justification out of the structure of wrongdoing and into the limbo of unanalyzed excuse.

Suppose, for example, that a detective loses patience with the criminal justice system because it has time and again released or acquitted his nemesis, a notorious hitman. The detective knows where the hitman lives and goes there, .357 Magnum in hand, to wreak vigilante justice. Approaching from the hitman's left, the detective kills him as he opens the door to his apartment. But the apartment is no longer the hitman's home. He sold the apartment the previous month, and ironically, the new owner turned out to be his next target. The hitman is holding his own gun in his right hand, where the detective cannot see it. At the moment the hitman is shot and killed, the target is standing right inside the door, frozen in terror, understandably convinced that he is about to die. Under the continuity view of criminal wrongdoing and the resulting genuine justification thesis, the detective is genuinely justified twice over.

misadventure, in self-defense, or while of unsound mind."); that it has encompassed a variety of more specific and quite distinct fault criteria, see *id*. at 548 ("'Express malice aforethought,' for example will be found to have four distinct meanings in different jurisdictions, even if occasional reversions to the ancient usage be ignored."); and that many defenses turn on its negation, see *id.* at 551 ("It has sometimes been said that every homicide is presumed to be with malice aforethought and that it devolves upon the prisoner to prove circumstances which will justify, excuse or mitigate the act.").

16. *See infra* Part IV.B.

17. *See infra* Part III.

18. *See infra* Part IV.C.

He has acted in defense of another[19] and in the execution of his law-enforcement duties.[20] The detective does not intend to kill justifiably, does not know that he is doing so, does not realize that he might be doing so, and is not behaving reasonably—but none of this matters to his justification defense. He is genuinely justified on the facts alone. Given that he is not only doing something wrong, but is also a deliberate law-breaker and a dangerous man, this is a hard conclusion to swallow.

This difficulty is not confined to the continuity view and the genuine justification thesis. Other analyses have the same troubling implication. But for these analyses, to say that the detective is excused from a murder conviction provides a way out. It opens the possibility that, even given the acquittal for murder implied by his mistake, the detective can be convicted of an attempt. Unlike an ordinary attempted murder, this crime involves a dead victim and not a surviving intended victim. Nevertheless, the defendant did intend to kill "unlawfully"—that is, unjustifiably—and failed to do so. This is the gist of an attempt.[21] The vague notion of an excuse facilitates this argument because it mitigates the awkwardness (which is acute, given that the victim is dead) of saying that the complete murder is genuinely justified but the attempt to commit the same murder is not justified.[22] We can say that he is guilty and that we excuse him from liability for the complete offense but not for the attempt. In contrast, the continuity view compels us either to embrace the awkwardness of imposing attempt liability unreservedly or to endorse the detective's acquittal. We might mitigate the difficulties of the genuine justification thesis if we say that the detective is legally justified, even if he is not morally justified. This is true so far as it goes. But for many, including me, to say this and nothing more introduces too much tension into the relationship between law and morality.

Given these difficulties, this is the place to ask why anyone would think the detective is justified to begin with. One reason is that to convict the rogue detective in spite of his mistake has its own troubling implications. His case suggests that we should require proof of a belief in justifying circumstances as well as proof of the circumstances themselves. What happens if we do this?

Suppose a woman is running in a park at dusk. She carries a small can of pepper spray to ward off any attack. Suddenly, a young man

19. MODEL PENAL CODE § 3.05 (1962) (providing that force may be used in defense of another if the actor would be justified in using force to defend himself against such a threat, the other person would be justified in using protective force, and the actor believes that his intervention is necessary for the protection of the other person).

20. *Id.* § 3.03 (providing that conduct is justifiable when it is required or authorized by the duties or functions of a public officer).

21. *See id.* § 5.01 cmt. 1 (explaining that attempt definitions "deal with situations where the actor has done all that he intends to do, but where the crime nevertheless has not been committed").

22. *See* SMITH, *supra* note 5, at 43 ("Can we sensibly say that, at one and the same time, a person is (i) justified in firing a gun at another with intent to kill him and (ii) guilty of an attempted murder?"). As it happens, attempt liability is not feasible for different reasons. *See infra* Part IV.C.

jumps out of the bushes and points a knife at her. She sprays him in the eyes with the pepper spray, which causes him extreme pain. It turns out that the knife is rubber, a toy, and the man meant to play a practical joke on a friend. In the fading light of day, he simply mistook the runner for his friend. The runner is charged with assault. She defends with a claim of self-defense. From an objective[23] point of view, the runner was never in danger, but from a subjective[24] point of view, she believed that she was. Under the continuity view of criminal wrongdoing, her honest and reasonable belief is sufficient to acquit her. To say that the runner has committed a crime but that we will excuse her from conviction and punishment seems an injustice—a magnanimous, but gratuitous, and entirely insulting gesture. To the contrary, she seems to have been completely justified in acting as she did. The practical joker was harmed, no doubt, but the runner did no wrong.

It is tempting to say that we should rewrite the law in such a way as to rule the runner's case within, and the detective's case out of, the class of genuinely justified acts. It might be possible to do this, but the attempt will leave us with two difficulties. First, this solution seems arbitrary. If we cannot give some principled rationale for rewriting the law in this way—that is, a rationale that is not entirely result-oriented—the law will lose a more than tolerable portion of its moral authority. The criminal law is nothing if not an interposition of reason in the world of violence and dishonesty. Second, without some consistent rationale to guide our redrafting, we risk doing more harm than good to the law, as a practical matter. The Model Penal Code's solution to this problem—to treat beliefs alone as dispositive—fails on both counts.[25]

Once set up by an elements analysis, an Aristotelian model of practical deliberation carries us to a principled revision of the law governing mistakes about justifying circumstances and a satisfactory solution to the problem of the mistakenly justified actor.[26] The law sometimes pursues incompatible objectives, such as the morally defensible treatment of the rogue detective and the runner, respectively. A specification of ends analysis reconciles these incompatible ends by a process of treating one as the means to the other, and then doing the same with the roles switch-

23. As used here, the term "objective" describes justification on the facts—from God's point of view, so to speak. This use of the term "objective" must be explicitly distinguished from "objective" as it is used in reference to criminal fault. Fault elements play a large role in my analysis, and it would be confusing to use "objective" for both of these meanings. In the latter use, it refers to non-intentional-states fault criteria such as negligence, extreme indifference, and depraved heart. I do not use the term "objective" to describe these fault criteria. The term "non-intentional" is more precise.

24. Similarly, the term "subjective" describes the basis of a mistake regarding justification: the beliefs of the defendant. This use of the term "subjective" must be explicitly distinguished from "subjective" as it is used in reference to criminal fault. In the latter use, it refers to intentional-states fault criteria such as purpose, knowledge, and advertent recklessness. I will use the term "intentional" to describe these fault criteria because it is more precise.

25. *See infra* Part III.C.

26. *See infra* Part IV.B.3.

ed. At each stage, the means is some particular mode of one of the conflicting ends, as determined by the other one. This is called the specification of ends. It is a kind of deliberation on ends—a distinctive feature of Aristotelian ethics—that occurs more or less implicitly in the ordinary processes of law making. Several lines of specification, each involving a different set of incompatible ends, often proceed in distinct lines toward making up the features of the law. If several lines of specification converge on one feature, then so much the better for purposes of an analysis or a defense of that feature.

The latter part of this Article offers a specification of ends analysis of the problem of the unknowingly justified actor, such as the rogue detective. Two lines of specification that run strongly through the criminal law as a whole—one line involving the conflicting ends of safety and autonomy and the other involving formality and fine-grainedness—exclude the mistakenly justified actor, such as the detective, from the category of genuine justification. This exclusion is accomplished by a formally and morally justified change in the elements of justification defenses—or, more accurately, by a change in the patterns of conjunction and disjunction that run continuously through offenses and justifications.

II. THE ELIMINATION OF EXCUSE

The overwhelming consensus is that the mistakenly unjustified actor is "excused." The best indication of this is that George Fletcher and Paul Robinson, whose disagreement over the case of the mistakenly justified actor is deep and longstanding, agree that the mistakenly unjustified actor is excused.[27] The difficulty is that the mistakenly unjustified actor is not excused in the sense of her being a non-responsible agent, such as an insane person or a minor—the most common meaning of "excuse" in American criminal law theory. This "excuse" apparently is closer in meaning to the English usage, which, I will argue, corresponds to the absence of criminal fault. My interpretation of "excuse" is by no means the prevailing view where mistakes in justification are concerned. But as others use it in this context, the term "excuse" is unanalyzed. For example, Heidi Hurd acknowledges the possibility that justification defenses can be understood as continuous with offenses in the way that I will describe in Part III, but she balks at the conclusion that such an actor is genuinely justified.[28] She writes: "[The law] ought, instead, to craft a special excuse doctrine to exonerate such a nonculpable person of the wrong that she has done."[29] Hurd acknowledges that "excuse" is used to refer to both claims of non-responsibility and missing fault elements of of-

27. *See* sources cited *supra* note 5.
28. Hurd, *supra* note 7, at 1564–65.
29. *Id*. at 1565.

fenses,[30] but the "excuse" she has in mind is apparently neither one of these things. What else it might be, she does not say.

This situation is not confined to mistakes regarding justification. The term "excuse" is used in several contexts in which it apparently does not refer to non-responsible agency, but in which it is not clear what its meaning could be, other than non-fault. As a preliminary to my main arguments, I will show that in these other contexts—duress, provocation, and mistake of fact—the idea of excuse collapses into either non-responsible agency or the absence of criminal fault. If "excuse" empties out in this way, then the mistakenly justified or unjustified actor is best understood to have acted without criminal fault. This understanding suggests, in turn, that justification defenses are continuous with offenses, and that mistakes regarding justification result in no crimes being committed.

A. The "Excuse" of Duress

Duress seems the strongest candidate for membership in a discrete category of excuse defenses. It never appears in offense definitions and, on first glance, it differs too much from its closest analogue—literal involuntariness—to be plausible as a claim of non-responsible agency. Take the paradigmatic case of the bank manager who is forced to take the bank's money out of the vault, thereby committing theft, and hand it over to a bank robber who threatens to kill the bank manager if he does not do so. To understand duress as a defense of non-responsible agency seemingly requires us to portray the bank manager as handing over the bank's money without a thought. This is too implausible. Crimes committed literally without a thought of doing otherwise must be, empirically speaking, few in number.

There are, however, several telling differences between duress on one hand and the other purported excuses, mistake and provocation, on the other. To begin with, the language we use in making these several arguments strongly suggests that the duress defense rests on the absence of responsible agency. We make mistakes and respond to provocation. We are placed under duress and are subjected to coercion. Mistake and provocation proclaim self-determination and self-respect. To argue that I acted under duress, however, is to claim that I suffered subjugation of the kind imposed by mental illness and extreme youth, and of the kind that undergirds the non-responsibility defenses of insanity and minority. The rest of the moral story of duress strengthens this analogy. The life of the actor under duress has a tragic dimension. He has committed a complete wrong that mars his life as a moral agent. His subjection to the ends and actions of another person is degrading. The fact of his doing wrong is another blemish on his life. The black marks of wrongdoing and

30. *Id*. at 1562–64.

degradation are the inevitable results of his own actions, yet they are no fault of his own. In this light it seems that we acquit the actor under duress as a matter of grace, out of compassion—and, here again, the acquittal of the mentally ill and the extremely young are close analogues.

One argument along these lines, however, needs to be avoided. It rests on the notion that the defendant who has committed a crime because of a threat has acted out of character, and that this is a reason to feel compassion and respond with grace. The problem is that we cannot, literally, act out of character because all of our acts go toward constituting our character. If a soldier runs from the enemy, this is not evidence that he is a coward. His running away makes him a coward, and constitutes a cowardly character. Likewise, the person who knuckles under to a threat and commits a crime has acted in a way that makes him a coward—even if just a little bit of one. If we acquit on grounds of character at all, it is because the actor under duress has acted *in* character. He has done as well as any of us would do under impossible circumstances. This success in meeting normative expectations constitutes good character; we are, in this sense, excused because of our good character.[31]

Success in meeting normative expectations is, as it turns out, a powerful theory of excuses beyond non-responsible agency (which, in my analysis, can be only claims of non-fault). But in the case of duress, we have to ask whether success in meeting the particular normative expectations at issue in the defense constitutes a good enough reason to acquit. The answer is no; this success is not why we acquit for duress. Showing a little bit of cowardice, instead of a lot of it, is not much of a reason to acquit.

This way of putting the argument is more glib than clear, so let me frame it in the same Aristotelian terms that Gardner used in advancing the idea of success in meeting normative expectations.[32] Gardner concedes that he is vague on the issue of normative expectations. He says the relevant expectations are those that concern our role as human beings in society. This does recall Aristotelian ethics, but sketchily. A more fully developed Aristotelian account specifies the relevant role as that of a deliberator on ends, whose actions are partially constitutive of society, and who is, therefore, responsible to society for the quality of his practical deliberations and the actions they produce.[33] The agent's deliberations on ends matter because this is where responsibility lies. For Aristotle, human beings are inherently social beings[34] and they have a distinctive *ergon*, a teleological purpose, of rational action.[35] To fail to reason well—and in particular to reason well with regard to the society of

31. *See* John Gardner, *The Gist of Excuses*, 1 BUFF. CRIM. L. REV. 575, 577–78 (1998).
32. *See id*. at 586 (identifying the normative expectations theory as "broadly Aristotelian").
33. *See* Huigens, *Virtue and Inculpation*, *supra* note 1, at 1447–49.
34. *Id*. at 1451–52.
35. *Id*. at 1449–50.

which one is partly constituted and partly constitutive[36]—is the salient failure of the wrongful actor in Aristotelian ethics. Given that it is concerned with this kind of responsibility, the adjudication of criminal fault is an assessment of a long-term process of assembling and maintaining the right kinds of values, concerns, and motivations.[37]

The defenses of mistake and provocation have an obvious and immediate bearing on this distinctively human role, and Gardner is right to describe the gist of these defenses as success in meeting normative expectations. Success in meeting the expectations relevant to mistake—to perceive and think accurately in all things—has a direct bearing on responsibility for acts. Whether the mistake is one of means-ends or a matter of deeper deliberations on ends, the mistake does not indicate fault if it does not indicate a deficiency in practical reasoning. As for provocation, to resort to lethal violence only in the most extreme circumstances demonstrates the actor's placing value on life, controlling his emotions, and heeding social norms—for the most part. These expectations are sufficiently nuanced to define degrees of homicide. In mistake and provocation, then, the normative expectations at issue are of the right weight and kind to be of concern to the criminal law. The defenses of mistake and provocation evaluate self-directed action that is fully reflective of the quality of the actor's deliberations on ends, and so count as defenses where these expectations are met.

Duress is different in this respect. The problem is not that duress blocks an inference to the quality of our deliberations on ends so that an assessment of normative expectations is impossible. This is an argument that Gardner warns against. I have repeated that warning above, and it is not my argument here. Our long-term deliberations on ends do indeed show up in our reactions to threats, as Gardner's normative expectations analysis supposes. The problem is that if the relevant normative expectations are expectations about the quality of our practical deliberations, then duress has little bearing on the relevant normative expectations. Responsibility lies in the effects our deliberations on ends have on our actions and, through them, on society. A person's reaction to threats has only a remote bearing on responsibility, because, beyond his immediate survival, the actor's ends are not at issue in his actions. His reaction to threats has only a remote bearing on fault because, beyond a brief inquiry into his choice among the limited available means to survival, assessing the quality of his practical deliberations is impossible. He has been compelled to pursue someone else's ends. It may take great courage to withstand sufficient duress, but even great courage in the face of an external threat tells us little about the quality of a person's deliberations on ends—particularly given the cramped, gerrymandered terms of

36. *Id*. at 1454–56.
37. *Id*. at 1460–62.

the duress defense.[38] In cases of duress, then, meeting normative expectations is not as good a reason to acquit as tragedy, compassion, and grace are. And if we acquit for these reasons, duress is a claim of non-responsibility.

B. The "Excuse" of Mistake of Fact

In English law and legal theory, a claim of mistake of fact is treated as an excuse.[39] Whereas the "excuse" of duress assimilates into the defenses of non-responsible agency such as insanity and minority, the "excuse" of mistake of fact assimilates into the defense of non-fault in the commission of an offense.

The Model Penal Code's principal innovation is section 2.02(1), which dictates that each material element of an offense shall be matched by a fault element—a "culpability" element in the terms of the Code.[40] A mistake of fact has the effect of negating one or more of these mandatory fault elements, resulting in a failure of proof on the part of the prosecution.[41] The scheme covers both honest and reasonable mistakes of fact. An honest mistake negates intentional states fault elements, because any belief to the contrary of a material element implies that not even so much as the risk that the material element might obtain has occurred to the actor. If the fault element of an offense is negligence—unreasonableness with respect to the material element—only a reasonable belief in the contrary of a material element will acquit.[42] The lesser showing on the part of the prosecution is reflected in a more demanding standard for the defense to meet.

In England, however, the change in mistake doctrine has been less thorough than in the Model Penal Code. Specifically, the principle of section 2.02(1) "is not and never has been a principle of English law."[43] English theory describes the matching of fault elements with material elements as correspondence, and the principle of 2.02(1) as rigid correspondence.[44] Free from the principle of rigid correspondence, English law

38. *See* MODEL PENAL CODE § 2.09 cmt. 1 (1962).

39. *See, e.g.*, JEREMY HORDER, EXCUSING CRIME 13 (2006); J.W. CECIL TURNER, KENNY'S OUTLINES OF CRIMINAL LAW 46 (Cambridge Univ. Press 1952) (1902) (stating conditions to be met "if mistake is to be held to excuse a harmful deed").

40. *See* Paul H. Robinson & Jane Grall, *Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond*, 35 STAN. L. REV. 681, 684 (1983) (describing section 2.02(1) as a "dramatic . . . innovation" with "far-reaching implications").

41. *See* MODEL PENAL CODE § 2.04 cmt. 1 ("Thus ignorance or mistake is a defense when it negatives the existence of a state of mind that is essential to the commission of an offense . . . .").

42. *Cf. id.* ("There is no justification, however, for requiring that ignorance or mistake be reasonable if the crime or the element of the crime involved requires acting purposely, or knowingly for its commission.").

43. John Gardner, *Rationality and the Rule of Law in Offences Against the Person*, 53 CAMBRIDGE L.J. 502, 509 (1994).

44. *Id*.

recognizes a number of constructive offenses.[45] For example, to cause death with an intent to cause grievous bodily harm constitutes murder.[46] Manslaughter includes not only reckless killings, but also killings done in the course of an illegal act, regardless of any fault specifically pertaining to death.[47] A separate homicide of causing death while driving dangerously is recognized.[48]

Looking beyond the Model Penal Code, however, the situation in America is not really very different in this respect. To begin with, the traditional felony murder formulation persists, a clear instance of constructive fault.[49] The Model Penal Code ties this doctrine to proof of an intentional-states fault element, but the commission of certain specified felonies creates a rebuttable presumption that this element has been proved.[50] This presumption itself is an instance of constructive fault. In some jurisdictions, the offense of misdemeanor manslaughter parallels felony murder.[51] A homicide by unlawful act is defined as manslaughter in some jurisdictions.[52] Homicide by dangerous driving, without proof of intent or negligence regarding death, is recognized in at least one state.[53] Constructive fault is alive and well in the New World.

The question these offenses raise in the present context is whether a mistake in the commission of a constructive fault offense is best described as an excuse instead of a failure of proof on fault. The neat scheme of the Model Penal Code, in which a discrete fault element is explicitly paired with a fact element, is not available. Given this fact, how else but as an excuse could the mistake defense be described?

45. *Id*.

46. *See* A.P. SIMESTER & G.R. SULLIVAN, CRIMINAL LAW: THEORY AND DOCTRINE 334–35 (2d ed. 2003).

47. *See id*. at 361; *see also* R. v. Kennedy, [2007] 1 A.C. 269 (H.L.) (appeal taken from Eng.) (discussing whether supplying the victim with a syringe full of heroin is a sufficient unlawful act); Alan Reed, *Unlawful Act Manslaughter and Affray*, CRIM. LAW., Nov.–Dec. 2006, at 1, 2 (describing the intent regarding the unlawful act that is required to support manslaughter liability).

48. *See* Road Traffic Act, 1988, c. 52, § 1 (Eng.).

49. *See, e.g.*, CAL. PENAL CODE § 189 (West 1999) ("All murder which is . . . committed in the perpetration of, or attempt to perpetrate, arson, rape, carjacking, robbery, burglary, mayhem, kidnapping, train wrecking, or any act punishable under Section 206, 286, 288, 288a, or 289, or any murder which is perpetrated by means of discharging a firearm from a motor vehicle, intentionally at another person outside of the vehicle with the intent to inflict death, is murder of the first degree."); MASS. GEN. LAWS ANN. ch. 265, § 1 (West 2000) ("Murder committed with deliberately premeditated malice aforethought, or with extreme atrocity or cruelty, or in the commission or attempted commission of a crime punishable with death or imprisonment for life, is murder in the first degree.").

50. *See* MODEL PENAL CODE § 210.2(1)(b) (1962).

51. *See* Bell v. State, 172 P.3d 622, 626–27 (Okla. Crim. App. 2007) (citing OKLA. STAT. ANN. tit. 21, § 711 (West 2001)) (misdemeanor driving while impaired supports manslaughter conviction); State v. Mickens, 123 S.W.3d 355, 370 (Tenn. Crim. App. 2003) (death in the execution of a dangerous misdemeanor supports manslaughter conviction "without any showing that the defendant intentionally, knowingly, recklessly, or even negligently caused the death").

52. *See, e.g.*, Commonwealth v. Fortes, 712 N.E.2d 104, 105 (1999).

53. *See, e.g.*, CAL. PENAL CODE § 192(c)(3) (West 2008) (manslaughter liability for death occurring in the course of automobile insurance fraud).

The answer is that, even in constructive fault offenses, the defense involves an offense definition. These offenses do not simply dispense with the deep and persistent doctrine that desert for legal punishment requires proof of criminal fault. The most common rationale for these offenses is not that they impose strict liability, but that the fault of the underlying crime or moral wrongdoing is attributed to the homicide. Gardner makes a straightforward attribution argument. He rejects the requirement of rigid correspondence, which allows him to rely on a general requirement of mens rea in the offense. This is said to be satisfied by the wrongdoing of the underlying offense or moral wrong, without regard to any one material element.[54]

In this light, however, constructive fault offenses offer no reason to think that mistake of fact should be considered an excuse instead of the negation of fault in an offense definition. The attribution rationale for constructive liability offenses places the so-called defense of excuse for mistake in the same place that rigid correspondence or the principle of section 2.02(1) puts it. It is a matter of fault, operating in the definition of offenses. It does not matter that discrete fault elements are not there to be negated. Fault attributed to one offense from another, or from morality to criminal law, is still fault in offense definition.

This fault cannot be negated by discrete beliefs to the contrary because it does not consist of discrete mental states. But as the example of negligence and reasonable belief shows, non-intentional fault criteria can be negated by other objective features of the offense as committed. Whether these features do negate non-intentional fault depends on what they tell us about meeting normative expectations. Given that fault is a matter of normative expectations, the substance of the excuse of mistake of fact, as Gardner describes it, is to be found in offense definition, and not in a separate defense.

C. The "Excuse" of Provocation

English law recognizes an excuse of diminished responsibility, which reduces murder to manslaughter on grounds of mental incapacity.[55] Because it has the same effect as a defense of provocation, the two arguments are regarded by some English practitioners as being interchangeable.[56] Noting this, Gardner points out the obvious: diminished responsibility is a claim of partial non-responsibility, which should not be considered an excuse at all. Provocation, however, is an excuse as Gardner describes it: an answer to the charge of murder that asserts the rational capacity of the accused and appeals to success in meeting normative expectations. In a case of provocation, the offender has exercised a

54. *See* Gardner, *supra* note 43, at 509.
55. Homicide Act, 1957, c. 11, §§ 2–3 (Eng.).
56. *See* Gardner, *supra* note 43, at 509–11.

sufficient degree of rational self-control to warrant capping his liability at manslaughter.

We can see why the excuse of provocation assimilates into non-fault in the commission of an offense if we notice another difference between these defenses. In the case of diminished responsibility, the reduction in the crime of conviction from murder to manslaughter is a matter of grace, as any defense of non-responsibility is. Given that "[t]he quality of mercy is not strain'd,"[57] there is no particular logic to this reduction. Nothing dictates manslaughter or some greater or lesser reduction in the grade of the offense as the consequence of a successful claim. Provocation is different, however. There is a logic to the reduction from murder to manslaughter because of the offender's adequate rational self-control. In the early codifications of criminal law, provocation or "heat of passion" negated the element of "malice aforethought" in murder. The offense fell to the level of manslaughter—the equivalent of a reckless killing—as a matter of offense definition.[58] The modern codified law of provocation follows the same logic, but more explicitly and precisely. This is the logic of offense elements, because provocation is not a defense at all. Provocation is an alternative definition of manslaughter.

The Model Penal Code manifestly conceptualizes provocation this way, because it places its version of provocation in the chapter on homicide as an alternative definition of manslaughter. A homicide that would otherwise be murder—a killing done purposely or knowingly—is manslaughter if it is also done "under the influence of extreme mental or emotional disturbance for which there is a reasonable explanation or excuse."[59] Oddly, almost all English and American commentary on the partial defense of extreme mental or emotional disturbance (EMED) shares the assumption that it is a partial defense.[60] The debate centers on whether it is a partial justification or a partial excuse. On the face of things, however, it is a mistake to consider EMED a defense of any kind, partial or full, excuse or justification.

To see EMED or provocation as a matter of offense definition, one has only to recognize that it is a non-intentional fault criterion. Non-intentional fault criteria, such as criminal negligence, are not marginal or honorific kinds of fault. On the contrary, they are paradigmatic. If criminal fault is a failure of practical reasoning or if non-fault is success in meeting normative expectations, we have no reason to think that inten-

57. WILLIAM SHAKESPEARE, THE MERCHANT OF VENICE act 4, sc. 1.

58. *See* Comber v. United States, 584 A.2d 26, 42–43 (D.C. Cir. 1990).

59. MODEL PENAL CODE § 210.3(1)(b) (1962).

60. *See, e.g.*, Joshua Dressler, *Provocation: Partial Justification or Partial Excuse?*, 51 MOD. L. REV. 467, 472 (1988) (provocation is a partial excuse); Alan Norrie, *The Structure of Provocation*, *in* 54 CURRENT LEGAL PROBLEMS 307, 338 (M.D.A. Freeman ed., 2001) (provocation is a partial justification). *But see* Stephen P. Garvey, *Passion's Puzzle*, 90 IOWA L. REV. 1677 (2006) (arguing that the provocation doctrine tries to distinguish killings committed in defiance of the law from those committed in a moment of culpable ignorance of law or weakness of will).

tional states are paradigmatic or constitutive of criminal fault. Broader inquiries into these questions are not only common,[61] they also are preferable if we seek a close match between our moral judgments and our legal judgments of wrongdoing. They facilitate a more fine-grained evaluation of the case than is permissible under the rule-like operation of proof and dis-proof of intentional mental states. This preference for non-intentional fault criteria seems problematic because they grant a large measure of discretion to the jury, presenting rule-of-law concerns.[62] Intentional states of mind are facts (or so the story goes), and finding them does not require the jury to make raw normative judgments such as "unreasonably" or "with extreme indifference to the value of human life." These are the reasons we prefer intentional-states fault criteria—but these are rule-of-law reasons that are exogenous to criminal fault itself.

Furthermore, a rigorous intentional-states approach to criminal fault has its own steep downside: the problem of the idiosyncratic believer. Suppose a karate expert decides to demonstrate his skill by aiming a lethal kick at a child's head—the point of the demonstration being that he has sufficient control to stop just short. He misses, killing the child. The appropriate charge under the Model Penal Code is murder on the theory that he was reckless "under circumstances manifesting an extreme indifference to the value of human life."[63] If he testifies truthfully that he honestly believed there was virtually no risk to the child's life and that this was a perfectly acceptable way to demonstrate his skill, he should be acquitted. He did not act recklessly because he did not "consciously disregard a substantial and unjustifiable risk" of death.[64] He did not believe the risk was substantial or unjustifiable.[65] This result will draw an outraged reaction from the public because it is so strongly at odds with our moral evaluation of the case.[66] Contrast this with the same case under a charge of murder "with malice aforethought" or "due to a depraved

61. I have in mind doctrines such as criminal negligence, unreasonable mistake in rape, transferred intent, felony murder, depraved heart or malice murder, strict liability, intoxication as a defense to recklessness, mistake regarding justifying circumstances, conditional intent, accomplice liability premised on negligence, and willful ignorance.

62. Criminal negligence is the most obvious example. Before the jury can decide whether the defendant has acted unreasonably, it must decide what reasonable conduct is.

63. MODEL PENAL CODE § 210.2(1)(b).

64. *See id.* § 2.02(2)(c) (defining recklessness as the conscious disregard of a substantial and unjustifiable risk).

65. Of course, this problem can be solved by treating "substantial and unjustifiable" as a mixed question of law and fact for the jury to determine, regardless of the beliefs of the defendant. But this option is one that advocates of the intentional-states conception of criminal fault reject. *See* Glanville Williams, *The Unresolved Problem of Recklessness*, 8 LEGAL STUD. 74, 77 (1988).

66. This was the reaction of the English public to the similar result in *Regina v. Morgan*, [1976] A.C. 182 (H.L.), in which the Law Lords held that an honest mistake regarding nonconsent was sufficient to acquit for rape. SANFORD KADISH & STEPHEN SCHULHOFER, CRIMINAL LAW AND ITS PROCESSES 322 (6th ed. 1996).

heart." The defendant would be convicted, in accord with our moral judgments.

What this example shows is that our rule-of-law concerns are in tension with our desire for morally defensible results in criminal cases. Formal norms remove discretion from legal decision-makers, give clear notice, and so on. But formal norms are inherently over- and underinclusive, and they produce counterintuitive outcomes as a result. Less formal legal norms, legal standards, are less over- and underinclusive, and produce fine-grained legal judgments that match our moral judgments. But the downside is that informality in legal norms presents rule-of-law problems, for example, lack of notice, vagueness, and excessive discretion. From this perspective, the formulation of fault elements in offense definitions constitutes a true legal art form: balancing formality and fine-grainedness in the formulation of legal rules and standards, to achieve the proper balance between these two conflicting ends. The principal materials of this art are intentional-states and non-intentional-states fault criteria.

Now look carefully at the Model Penal Code's homicide definitions. One kind of manslaughter uses an intentional-states fault criterion: advertent recklessness. The other kind of manslaughter uses a mixed intentional and non-intentional fault criterion: purpose or knowledge combined with EMED. It is important to note that the non-intentional part of this fault criterion includes not only the subjective[67] "disturbance" but also the absence of a "reasonable explanation or excuse" for the EMED. Historically, a claim of provocation had to be premised on one of several objective circumstances in an exclusive list—for example, sudden discovery of adultery or mutual combat.[68] The demand for a reasonable explanation or excuse is the Model Penal Code's equivalent of these non-intentional criteria. Arguably, then, the absence of a reasonable explanation or excuse—the non-intentional dimension—should predominate in the application of this definition of manslaughter. Indeed, an excessive emphasis on the subjective dimension of EMED has been notably bemoaned, albeit in confused theoretical terms.[69]

The Model Penal Code's definitions of murder bolster this analysis of EMED manslaughter because murder, too, is defined in the alternative, using intentional-states and mixed intentional, non-intentional fault criteria, respectively. A murder is a killing done purposely or knowing-

67. A subjective fault element is not necessarily an intentional-states fault element. All intentional states are subjective, but there are subjective states—such as an emotional or mental "disturbance"—that are not intentional states of mind.

68. *See* Girouard v. State, 583 A.2d 718, 721 (1991) (stating that provocation is traditionally limited to cases of extreme assault or battery, mutual combat, illegal arrest, injury to or abuse of a close relative, and sudden discovery of adultery).

69. *See* Victoria Nourse, *Passion's Progress: Modern Law Reform and the Provocation Defense*, 106 YALE L.J. 1331, 1394, 1440 (1997) (describing ill effects of the Model Penal Code's "extreme mental or emotional disturbance" formulation and reconceptualizing provocation as a "warranted excuse").

ly.[70] As an alternative, a murder is a killing done recklessly with extreme indifference to the value of human life: an intentional-states and a non-intentional-states fault criterion.[71] This set of alternative definitions of murder not only resembles the alternative definitions of manslaughter—in fact, the four definitions are symmetrically complementary. A reckless killing that would otherwise be manslaughter is promoted to murder based on the non-intentional criterion of an extreme indifference to human life. A purposeful or knowing killing that would otherwise be murder is demoted to manslaughter based on the non-intentional fault criterion of a mental or emotional disturbance for which there is no reasonable explanation or excuse. The defense of provocation assimilates into the offense definition of homicide. No discrete "partial excuse" of provocation remains.

Given that duress assimilates into non-responsible agency defenses, and that provocation, along with mistake of fact, falls into the category of disproving the fault elements of offenses, there is no discrete category of "excuse" defenses. We might, as Hurd suggests, improvise a special category of excuse to cover mistakes regarding justification, but in this light this would seem to be an especially ad hoc and inelegant analysis of these cases. We will do better to analyze the justification defenses as being continuous with offenses. Once we separate them conceptually from "excuse," we can see that fault elements operate in justification defenses just as they do in offenses. The resulting arrangement of elements is the core of the continuity view of the justification defenses.

III. TWO VERSIONS OF THE CLOSURE VIEW OF JUSTIFICATION DEFENSES

Is there a substantive difference between criminal law offenses and justification defenses, or is the arrangement of their respective elements a matter of custom and convenience? Two of the great criminal law theorists of the last century thought there was no difference. H.L.A. Hart took a generally pragmatic approach to the details of criminal law doctrine,[72] believing that any defense to homicide amounted to a claim that "some element, negative or positive, required in the full definition of criminal homicide (murder or manslaughter) was lacking."[73] Glanville Williams was more pointed and strident on the question, though equally resistant to theoretical explanations.

70. MODEL PENAL CODE § 210.2(1)(a) (1962).

71. *Id.* § 210.2(1)(b).

72. *See* H.L.A. HART, *Legal Responsibility and Excuses*, *in* PUNISHMENT AND RESPONSIBILITY 28, 31 (1968) (describing non-excuse, sanity, non-duress, involuntariness, and provocation, without distinction, as the absence of the "'subjective element' or 'mens rea'").

73. H.L.A. HART, *Prolegomenon to the Principles of Punishment*, *in* PUNISHMENT AND RESPONSIBILITY, *supra* note 72, at 1, 13.

> The general principle is that a mens rea word in a statutory offence governs all the elements of the offence in the absence of words to the contrary, and this principle should be applied to exceptions to offences, without quibbling about whether or not they are an outgrowth of the definition.[74]

Williams's position in this regard is known as the "negative elements" view of the justification defenses.[75]

In this Part, I will consider some arguments against the negative elements view and in favor of the notion that there are fundamental differences between offenses and justification defenses. It seems to me that the simplest argument for that conclusion has been overlooked, and that it is obvious once we lay out the elements of offenses and justification defenses on a table that shows not only the elements but also the patterns of conjunction and disjunction between them. However, I will go on to argue that the negative elements view can be salvaged with a modest adjustment. This adjustment gives us the continuity view of justification defenses.

A. The Negative Elements View of the Justification Defenses

Consider this novel account of a familiar picture:

> Combining the closure view of wrongdoing with the remainder view of responsibility, a simple and appealing framework emerges according to which (a) justifications are denials of wrongdoing, (b) excuses are denials of responsibility, and (c) wrongdoing and responsibility combine (jointly exhaustively and mutually exclusively) to yield adverse normative consequences such as liability to be punished for one's wrongs.[76]

The familiar picture is the "a, b, c" sequence. It is commonly used in American law schools to make intuitive distinctions between, say, self-defense and insanity. One who defends himself from another person who threatens him has a right to do so, and seems to have done no wrong.[77] In contrast, an insane actor's argument is that he is not a responsible agent, and is therefore not a fair candidate for condemnation

74. Glanville Williams, *Offences and Defences*, 2 LEGAL STUD. 233, 240 (1982).

75. *See* Andrew Simester, *Mistakes in Defense*, 12 OXFORD J. LEGAL STUD. 295, 305–09 (1992) (making a doctrinal and normative case against subsuming justification defenses into offense definitions as "negative elements").

76. Gardner, *Defenses*, *supra* note 9, at 9.

77. *See* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 222 (2006) ("[J]ustified conduct is conduct that under ordinary circumstances is criminal, but which under the special circumstances encompassed by the justification defense is not wrongful and is even, perhaps, affirmatively desirable.").

and incarceration.[78] Criminal liability requires a complete wrong done by a responsible agent, or unjust punishment will ensue.[79]

This familiar picture makes sense of the most common distribution of burdens of persuasion on the defenses. The prosecution must disprove self-defense because it must prove wrongdoing, and justification negates the wrong.[80] In contrast, a defendant who raises an insanity defense has conceded that he violated a prohibition and that he has no justification for doing so, yet invokes a rule that requires us to acquit him. In light of his concessions—in light of the complete wrong he has done—it is only fair to require him to persuade us that he deserves this extraordinary treatment.[81]

The unfamiliar terms "closure" and "remainder" reflect this standard picture. The closure view of wrongdoing is the idea that criminal wrongdoing is not complete unless and until justification has been excluded from the picture. Proof that the accused has violated a criminal prohibition opens up the possibility of criminal liability, but the analysis is not complete—the account is not closed, so to speak—unless and until the exceptions to that prohibition have been considered. The remainder view of responsibility is a practical consequence of the closure view of wrongdoing. When all arguments concerning the criminal act have been absorbed into the concept of wrongdoing, nothing is left of excuse except the idea of non-responsibility. As I noted above, mistake of fact used to be considered an excuse. Strange as it sounds, defenses such as self-defense were once commonly referred to as excuses.[82] But if the justification defenses are treated instead as exceptions to the prohibition, they become as much matters of wrongdoing as offenses are. The only defense still independent of the question of wrongdoing, and as a result the only one meaningfully called an excuse, is a claim of non-responsible agency, such as insanity or minority.

The standard picture of justification defenses as a matter of wrongdoing and excuses as a matter of non-responsibility is relatively new; and, notwithstanding the support for it offered by Hart and Williams, it is open to serious debate. Primarily in England, where criminal law is less

78. *See* Sanford H. Kadish, *Excusing Crime*, 75 CAL. L. REV. 257, 264 (1987) ("To blame a person is to express a moral criticism, and if the person's action does not deserve criticism, blaming him is a kind of falsehood and is, to the extent the person is injured by being blamed, unjust to him.").

79. *See* SANFORD H. KADISH ET AL., CRIMINAL LAW AND ITS PROCESSES: CASES AND MATERIALS 737 (8th ed. 2007) ("Thus, claims of self-defense and insanity both suggest reasons to bar conviction even when it has been clearly proved that the defendant killed someone intentionally.").

80. *Cf.* Martin v. Ohio, 480 U.S. 228, 236 (1987) ("We are aware that all but two of the States, Ohio and South Carolina, have abandoned the common-law rule and require the prosecution to prove the absence of self-defense when it is properly raised by the defendant.").

81. *Cf.* JOSHUA DRESSLER, CASES AND MATERIALS ON CRIMINAL LAW 619 (4th ed. 2007) (attributing the majority rule imposing the burden of persuasion for insanity on the defendant to anger over John Hinckley's acquittal for his attempt to kill Ronald Reagan).

82. *See, e.g.*, Joseph Raz, *Legal Principles and the Limits of Laws*, 81 YALE L.J. 823, 832 (1972) ("The law instituting self-defense or mistake of fact as an excuse always overrides the law prohibiting assault.").

codified and the structure of criminal liability is more contested as a result, the wrongdoing and non-responsibility picture of criminal liability has its detractors. Among these are, most notably, Andrew Simester and John Gardner. They share the view that offenses and defenses, including justifications, are conceptually distinct from one another.[83] Accordingly, they reject the closure view of criminal wrongdoing.

Simester describes a sequence of cases that, in his view, obscured the difference between an offense and a justification defense in English law, and gave the negative elements view of justification defenses unwarranted support.[84] In *Regina v. Morgan*,[85] the Law Lords held that a mistake about the nonconsent of the victim in a rape case need only be an honest mistake—no matter how unreasonable—in order to acquit.[86] In *Regina v. Williams*,[87] this holding was extended to the element of "unlawfully" in the definition of assault. As with nonconsent and the other elements of offenses, a mistake about the lawfulness of one's actions need only be honest, not honest and reasonable, in order to acquit.[88] In *Beckford v. Regina*,[89] the Lords took *Williams* to its logical conclusion, as they saw it. If "unlawfully" in the definition of assault offers a justification to the defendant who can negate it—to act with justification is to act lawfully, not unlawfully—then just as a belief in lawfulness need only be honest, a belief in other justifying circumstances need only be honest, no matter how unreasonable.[90] In *Beckford*, a police officer shot and killed a suspect who had his hands raised, apparently surrendering to the officer. The officer claimed that he honestly believed that the suspect still presented a deadly threat to him. His conviction was overturned. The jury had been instructed that this belief had to be both honest and reasonable. The Lords held that his belief in the deadly threat need only be honest to acquit.[91]

At this point, let me introduce a system of notation to describe the progression from *Morgan* to *Williams* to *Beckford*. It will help in seeing Simester's argument; but, more important, it can be extended and elaborated in ways that will help me present my arguments about offenses and justification defenses. Take "HB" to represent honest belief, "RB" to represent reasonable belief, "OE" to represent offense elements, and "JE" to represent justification elements. In addition, let "~" stand for negation, or "not"; let "∨" stand for disjunction, or "or"; and let "•"

83. *See* Gardner, *Defenses*, *supra* note 9, at 254; Simester, *supra* note 75, at 305–09 (making a doctrinal and normative case against subsuming justification defenses into offense definitions as "negative elements").

84. Simester, *supra* note 75, at 298–300.

85. [1976] A.C. 182 (H.L).

86. *Id*. at 214.

87. [1987] 3 All E.R. 411.

88. *Id*. at 413–14.

89. [1988] 1 A.C. 131 (P.C. 1987).

90. *Id*. at 144.

91. *Id*. at 141.

stand for a conjunction, or "and." Finally, in describing the defendant's evidentiary burdens—both to present evidence and to raise a reasonable doubt—let us use lowercase letters. Later, for the prosecution's burden to persuade beyond a reasonable doubt, we will use uppercase letters. With this notation, I will construct tables describing the elements the prosecution and defense must show, as well as the connections between these elements. Please take note that the tables do not purport to assign evidentiary burdens as a normative matter. They are entirely descriptive, laying out the proof that each side must offer if it is to prevail under existing legal rules and under the various legal rules at issue in arguments about the nature of criminal wrongdoing and the justification defenses.

Morgan's holding can be represented this way, in terms of the defendant's evidentiary burden to make a prima facie case or to raise a reasonable doubt about the offense element, "oe," of nonconsent. The several offense elements are represented here as "oe^1, . . . oe^n."

hb	∨	$\sim oe^1$
	∨	
hb	∨	$\sim oe^n$

The defendant might, of course, show that the offense element of nonconsent was absent, i.e., that the victim did consent. *Morgan* concerned the other alternative, a showing that he had an honest belief, "hb," to this effect, holding that the left-hand alternative need not be "rb," or a reasonable belief. Of course, either showing on either horizontal axis will be sufficient to acquit, making the horizontal and vertical connectors "∨."

In *Williams*, an assault case, the Lords recognized that "unlawfully" is a (non-) justification element,[92] but they treated it as being indistinguishable from the several offense elements of assault. (The defendant's showing that he did not act unlawfully—that is, that he is not unjustified—is represented as "~~je." This is, of course, equivalent to "je" which is how other justification elements—such as necessity to use force—are represented.)

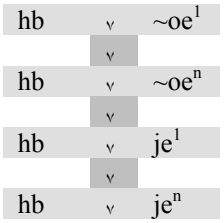
hb	∨	$\sim oe^1$
	∨	
hb	∨	$\sim oe^n$
	∨	
hb	∨	~~je

Thus, the defendant might show that he did not cause bodily injury or he had an honest belief that he did not cause bodily injury, or that he did not use a deadly weapon or had an honest belief that he did not use a deadly weapon. Likewise, he might show that he did not act unlawfully—i.e., that he acted with justification—or that he honestly believed he

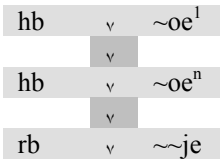
92. *Williams*, 3 All E.R. at 411.

was acting lawfully. Any of the six alternatives would suffice to acquit him, of course, because the prosecution is required to prove all the fault and fact elements of the offense charged.

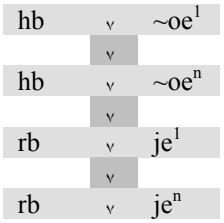
In *Beckford*, the Lords held that an honest belief in any of the several elements of the justification defense of self-defense would suffice to acquit the defendant.[93] In our terms,



Simester's complaint against *Williams* and *Beckford* is that the Lords ignored longstanding authority—which *Morgan* did not undermine—that a mistake concerning offense elements need only be honest, but that a mistake concerning justification elements must be reasonable.[94] If the *Williams* court recognized—as it did—that "unlawfully" offers a justification, then the elements of assault were properly this:



Where the justification defense is separately defined, as self-defense is, it is even more clear that a mistake concerning justification elements must be reasonable:



This, then, should have been the holding in *Beckford*.

Simester notes that Glanville Williams's "negative elements" view of justification defenses supports *Beckford*. "In support of such an approach, Glanville Williams has asked why, if a negative fact like non-consent can be an element of the *actus reus* of rape, should not the negative of all 'so-called' defences rank among such elements?"[95] As noted, Simester objects that this overrides the traditional rule that a reasonable

93. *Beckford*, 1 A.C. at 144.
94. Simester, *supra* note 75, at 299–300.
95. *Id*. at 304 (citing Williams, *supra* note 74, at 242).

mistake is required to acquit where the mistake concerns a justification. But Simester never gets at the real weakness of Williams's position.

We can express what Williams apparently had in mind in terms of an elements analysis with the addition of a few more terms that will allow us to express the prosecution's burden of persuasion. As I mentioned above, I use uppercase letters to describe the elements of the prosecution's case, which must be proved beyond a reasonable doubt. At a minimum, the prosecution must show that the offense elements obtain: $OE^1, \ldots OE^n$. Usually, the prosecution also has to show fault in conjunction with the offense elements,[96] and the fault criterion usually is an intentional state of mind. The Model Penal Code defines three intentional-states fault criteria: purpose, knowledge, and recklessness.[97] It also provides a default rule to govern where, as is usually the case, the legislature has not specified a fault element for a given material element. This rule is that purpose, knowledge, or recklessness will suffice, hence "$\geq$R."[98] This level of fault, $\geq$R, is rebutted by the defendant's honest belief, hb, that the offense elements do not obtain. A defendant does not have an intentional state of mind regarding a circumstance or result element if he has an honest belief to the contrary.[99]

Finally, I use the Greek letter êta, "η," to indicate a normative implication, or "therefore." The truth of the defendant's case obviously does not imply the truth of the prosecution's case as a matter of fact. Just the opposite is true, of course, and so we cannot use the symbol "$\supset$," which stands for implication between propositions. But the truth about what the defendant must prove does imply the truth about what the prosecution must prove, within the normative system of criminal law. Systems of formal logic for norms do exist, but they are unnecessarily complex for our limited purposes.[100] The gist of the tables I will construct is that, if the defendant's case is as described, the prosecution's case must be as described, as an implication within the normative system of criminal law. The symbol "η" refers to this kind of implication.

Under this system of notation, Williams's position is this:

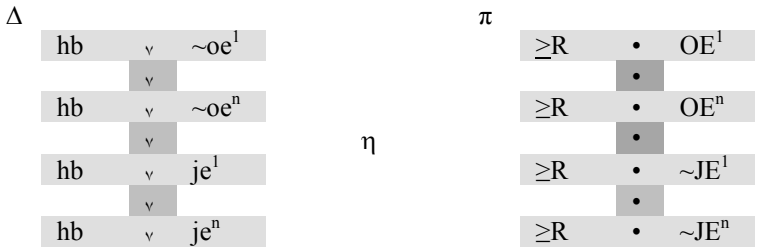
96. *See, e.g.*, MODEL PENAL CODE § 2.02(1) (1962) (requiring proof of fault for each fact element).

97. *Id.* § 2.02(2).

98. *Id.* § 2.02(3).

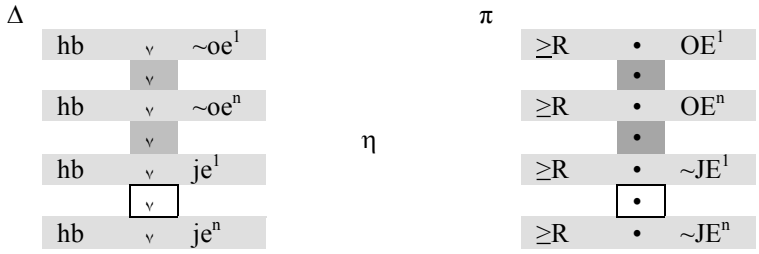
99. *See id.* § 2.04 cmt. 1.

100. *See, e.g.*, Alan Ross Anderson & Omar Khayyam Moore, *The Formal Analysis of Normative Concepts*, 22 AM. SOC. REV. 9 (1956).



The defendant's burden to raise a reasonable doubt about whether an offense element obtains *or* to raise a reasonable doubt about whether he has an honest belief obtains, implies that the prosecution must show beyond a reasonable doubt that the offense element obtains *and* that the defendant has an intentional state of mind regarding the offense element.[101] To put this another way, the burden on the prosecution to prove ≥R • OE implies that the defendant must show hb ∨ ~oe. That is, he must provide evidence that either will raise a reasonable doubt about whether he committed the offense, ~oe, or will constitute a prima facie case of mistake of fact, hb, that, if unrebutted, will constitute a reasonable doubt about the converse fault elements of the prosecution's case, ≥R. Finally, Williams proposes that the absence of justification elements can be treated just as offense elements are. That is, the prosecution must prove ≥R • ~JE, and the defense must prove hb ∨ je.

Once we lay it out on an elements table, we can see that the negative elements view is clearly wrong on the vertical axis. The implication of Williams's argument is that, as with offense elements, the prosecution would have to prove every aspect of the defendant's not being justified. In elements analysis terms, where a defendant has made a prima facie case of self-defense, the prosecution would have to show the non-existence of a threat with at least recklessness as to the non-existence of a threat, *and* non-necessity with at least recklessness as to non-necessity, *and* non-imminence of the threat with at least recklessness as to the non-imminence, and so on through the negation of each justification element and its associated fault element. Here is the negative elements view, with this aspect highlighted:



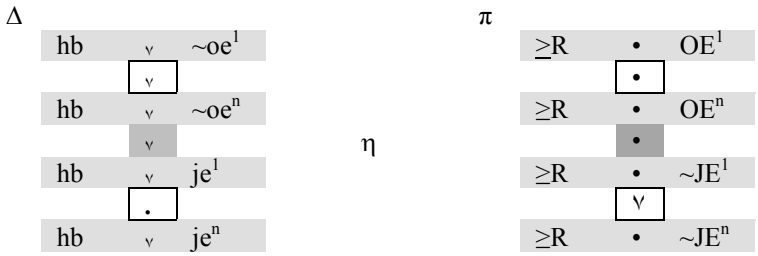
101. This is an application of DeMorgan's Theorem. The denial of two propositions taken in conjunction is the equivalent of the negation of each of these propositions taken disjunctively. To say that this object is not a red apple is to say either that it is not red or that it is not an apple.

The problem lies with these vertical connectors. In all jurisdictions, the fact that any one of these justification elements is missing, and that the defendant was at fault in this respect, is enough to deprive her of the defense.[102]

For example, suppose someone attacks me with a gun that obviously is made of plastic. I do not know that it is a toy gun—I know there are real, lethal guns that are made of plastic—but I know it is very, very likely to be a toy gun. In any event, I shoot this "attacker" and claim self-defense. If, as in this case, I am aware of a substantial risk that it is not necessary to use deadly force against an attacker wielding what is probably a toy gun, I ought not to succeed in a claim of self-defense. And surely this showing on the question of necessity is enough to convict me. The prosecution does not have to prove, in addition to this, non-threat with at least recklessness as to non-threat, *and* non-imminence with at least recklessness as to non-imminence, *and* so on. Proof of one non-justification element, along with fault as to that element, will suffice. The gunman might actually have posed an imminent threat, and I might have honestly believed this. It was clear that, if all else failed, he would throw his plastic gun at me. But this cannot be a sufficient reason to acquit me. I ought to lose on my claim of self-defense if I knew it probably was not necessary to use deadly force to adequately protect myself, regardless of what else I believed.

A justification defense cannot be a mere extension of the offense, as the negative elements view has it, because the vertical connectors flip as the proof moves from offense to justification defense. On the defense side, offense elements are vertically disjunctive and justification elements are vertically conjunctive. On the prosecution side, offense elements are vertically conjunctive and non-justification elements are vertically disjunctive.

102. *See, e.g.*, People v. Dillard, 745 N.E.2d 185 (Ill. App. Ct. 2001) (holding that if the state negates any of the elements of self-defense, the defendant's claim of self-defense must fail); People v. Dunlap, 734 N.E.2d 973 (Ill. App. Ct. 2000) ("A defendant, in order to raise a claim of self-defense, must present evidence supporting each of the following elements . . . ."); State v. Smith, 472 A.2d 948 (Me. 1984) (holding that is was not error to give the following jury instructions: "The State has the burden but the State need not eliminate all of the elements of self-defense beyond a reasonable doubt. The State need only eliminate one or more of them in order to overcome the defense of self-defense or the justification defense."); People v. Acquista, 837 N.Y.S.2d 309, 310 (N.Y. App. Div. 2007) ("Indeed, the defendant's own testimony concerning his near-fatal stabbing of the complainant, as well as other evidence at the trial, plainly disproved beyond a reasonable doubt at least two elements of the defendant's justification defense."); Juarez v. State, No. D3-04-00248-CR, 2005 WL 1939956, at *3 (Tex. App. Aug. 11, 2005) ("An appellant must raise some evidence on *each* element of the defense in order to warrant a self-defense instruction.").



I will elaborate on this difference immediately below and will defend it and draw some implications from it further on, but for now it is enough to note one apparent implication. This elements analysis supports Simester's and Gardner's view that there is a genuine difference between offense and defense in criminal law.

This difference is reflected in the normative features of the justification defenses. Gardner points out that there are moral consequences for wrongdoing short of punishment, such as "a duty to show regret, to apologize, to make restitution, to provide reparation, and so on."[103] He notes that one such consequence is of particular importance to law: the duty simply to offer a justification for one's wrongdoing. Gardner finds this norm instantiated in the legal duty to meet a burden of production, if not a burden of persuasion, on the question of justification.[104] Simester might add that this is why a mistake about a justification must be reasonable, whereas a mistake about an offense need only be honest.[105] And Gardner's point about the moral consequences of wrongdoing goes some way toward explaining why justification elements are vertically conjunctive on the defense side, whereas offense elements are vertically disjunctive. All three of these burdens on the defendant—each of them greater in the proof of a justification defense than in the disproof of an offense—seem to reflect the fact that he has committed a complete wrong, even if he was justified in doing so.

As persuasive as these formal and normative points might be, however, they run contrary to my thesis. Whereas they point to discontinuity between offenses and justification defenses, I will argue that offense and defense are better described as continuous.

B. The Continuity View of the Justification Defenses

I will argue for a different closure view from the negative elements view advanced by Williams (and, less explicitly, Hart). I call this the continuity view. In fairness, I must make one point clear before we go any

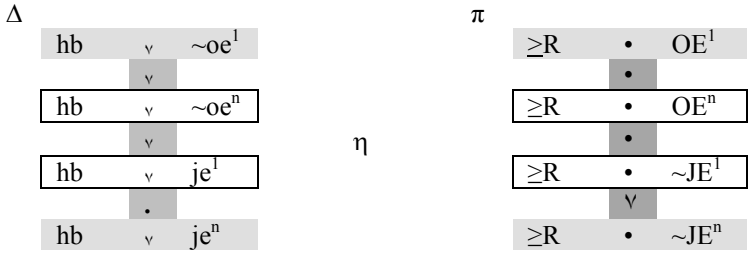
103. Gardner, *Defenses*, *supra* note 9, at 256.

104. *Id*.

105. Simester, *supra* note 75, at 308–09 (arguing that moral differences between offense and defense support requiring proof of honest mistake in the former, but reasonable mistake in the latter, to acquit).

further. The superiority of the continuity view to the negative elements view is obvious in light of the extant case law on the distinguishing point—the flip in connectors on the vertical axis as the proof moves from offense to justification defense. Williams surely knew this, so it is possible that what Williams had in mind when he advanced the negative elements view is what I call the continuity view. From this point on, however, I will deliberately ignore this possibility, for two reasons. First, the defect in the negative elements view, as obvious as it might be, seems not to have been noted before, or, if it has been noted, its significance has been underappreciated. Simester, for example, does not mention it in his doctrinal case against the negative elements view. Second, the negative elements view serves as a helpful foil in elaborating and defending the continuity view.

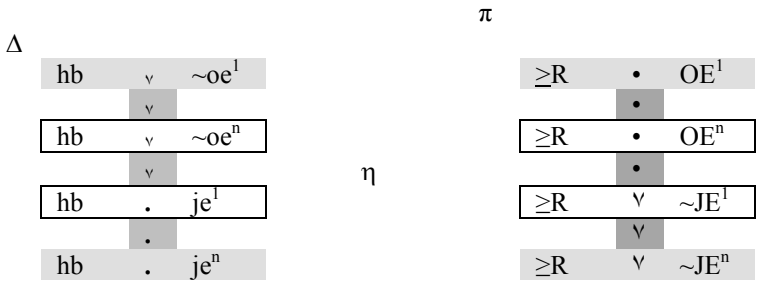
There are two ways in which a closure view might be wrong. It might be wrong about the connectors on the horizontal axes, and it might be wrong about the connectors on the vertical axis. We have already seen that the negative elements view is wrong on the vertical axis, and that the continuity view is defined by its correcting this defect. There is no such difference on the horizontal axes. On both the continuity and the negative elements views, the connectors between fault elements and justification elements on the horizontal axes are disjunctive, or "∨," on the defense side, just as they are for fault elements and offense elements. On the prosecution side, of course, these horizontal connectors are conjunctive, or "•." Here is the continuity view, highlighting the similarity of these horizontal axes:



In terms of an elements analysis, the negative elements view overstates the continuity between offenses and justification defenses in supposing that the vertical connectors are the same for offense and defense. Ironically, this discontinuity is what distinguishes the continuity view from the negative elements view. The continuity in the continuity view refers to the pattern of connectors on the *horizontal* axes: disjunctive throughout for the defense; conjunctive throughout for the prosecution.

From a competing point of view, this continuity on the horizontal axes is precisely what makes the continuity view wrong. Under what is known as the *Dadson* view, the connectors between fault elements and justification elements on the defense side are conjunctive, or "•," and

those on the prosecution side are disjunctive, or "∨."[106] This is different from the connectors between fault elements and offense elements—a point that is seldom, if ever, fully appreciated. Here is the *Dadson* view, highlighting these contrasting horizontal axes:



Near the end of this Article, I will argue that the *Dadson* view is incorrect. For the time being, let us assume that this is so. I will leave the horizontal connectors as the closure view has them—in a continuous pattern through offense and defense—and look at the implications of the continuity view.

Consider the case of the runner I described in the introduction. As a matter of fact, she is never in danger. The prosecution could show ~JE all the way down. However, on the *horizontal* axes, the prosecution must show that she is at least reckless, ≥R, with respect to one of these non-justification elements. The runner prevails in this case because the prosecution cannot do this. It is enough for an acquittal if the runner shows that she honestly believed she was in danger in every respect that the definition of self-defense prescribes. And this is what we have. The prosecution can show non-justification in the runner's case, ~JE, but it cannot show fault, ≥R.

On the continuity view, the runner is not only properly acquitted, she is genuinely justified. Consider the classic case of mistake of fact. I take and intend to keep an umbrella that does not belong to me, but I believe that it does belong to me. To convict me of theft, the prosecution must prove that the umbrella is the property of another person and must also prove that I was at least reckless concerning the umbrella's being the property of another person. But it never occurred to me that the umbrella belonged to someone else, and so I neither hoped, nor knew, nor considered a risk that it did. On the facts as given, we say that no crime has been committed. An essential element of the offense—a fault element—is missing, and the prosecution has failed to prove its case. The umbrella thief is not excused in either of the two senses in which that term is used: he does not lack moral agency, and he does not appeal to any conception

106. *See* Christopher, *supra* note 5, at 229 (stating that "justified force requires belief in, or knowledge of, the presence of justificatory circumstances") (citing R. v. Dadson, [1850] 4 Cox C.C. 358 (Ir.)).

of excuse purportedly distinct from such a plea of non-responsibility. He does not need to do so. There is no criminal wrongdoing to begin with because the prosecution has failed to prove at least one constituent element of the offense: fault with respect to the umbrella's being the property of another person.

On the continuity view of justification defenses, the runner is essentially in the same position as the umbrella thief. There is a difference: in her case, all the material and fault elements of the offense, assault, are easily proved. But just as in the case of the umbrella thief, there is no crime because the prosecution has failed to prove a constituent element of its case: either fault with respect to non-threat or fault with respect to non-necessity, or both. Given this failure of proof, the runner can be described only as genuinely justified.

The difficulty, of course, is that the rogue detective is also genuinely justified on the continuity view. He can gain an acquittal by showing either the factual elements of the justification or an honest belief in those elements, or even a combination of the two, provided he can show at least one of each pair. This means that the facts of his case are sufficient to acquit him, regardless of the state of his beliefs. This strikes most observers as the wrong result, which explains the attractions of the *Dadson* view. The facts of the rogue detective's case are not sufficient to acquit him. He must also believe that they exist, which is something he cannot show. If we require him to show justifying facts and a belief in them, he will be convicted, but so will the runner, who has the correct belief but not the necessary justifying facts. We are caught in a dilemma.

C. Continuity and the Model Penal Code

The continuity view of criminal wrongdoing is essentially the view adopted by the Model Penal Code.[107] The Code's principal innovations occurred in the area of criminal fault, or, as it is known in the Code, culpability. Section 2.02(1) provides, "[A] person is not guilty of an offense unless he acted purposely, knowingly, recklessly or negligently, as the law may require, with respect to each material element of the offense."[108] The effect of this provision is to set up a double column of elements of the kind that I have used to describe the negative elements and continuity views of justification defenses. On the left hand we have fault elements, and on the right, factual elements. The prosecution must prove each pair. My application of this scheme to justification defenses can be contested, but it is not novel. The Code itself extends the principle of 2.02(1) to the justification defenses in its definition of material elements,

107. I use the word "essentially" advisedly here. This is not the Code's actual position; it is the position the Code adopts before it adopts a defective second-best solution to a dilemma presented by the continuity view and another item on the drafters' agenda.

108. MODEL PENAL CODE § 2.02(1) (1962).

which expressly includes elements of justifications and excuses, as well as of offenses.[109] This definition supports not only my claim that justification defenses are continuous with offenses with respect to the proof of fault, but also my argument that an excuse that is not a claim of non-responsibility is a claim of non-fault. I will not, however, rely on these "definitional stop" arguments.[110]

The better argument for extending the double column scheme to justification defenses has to do with a third subsection of the same statute: 2.02(3). Legislatures do not, and are not expected to, comply with 2.02(1). They are expected, instead, to do what they have always done—to drop fault criteria into offense definitions in plausible but unsystematic ways. Section 2.02(3) is the better of two somewhat inconsistent default rules[111] provided for the purpose of bringing actual offense definitions into compliance with 2.02(1) for purposes of charging documents and jury instructions. The signal feature of this default rule is its preference for intentional-states fault criteria. Proof of purpose, knowledge, or recklessness will suffice to establish fault for unmatched material elements, whereas criminal negligence will not.[112]

This exclusion of negligence from the default rule on fault has an important implication for the Code's treatment of mistake. If intentional-states fault is required to convict for most offenses, then most offenses are not committed if the accused had an honest belief to the contrary of some material element.[113] If it never occurred to the accused that the circumstances or the result of his conduct would be one of those proscribed in the offense alleged, then he cannot have been reckless, knowing, or purposeful about that circumstance or result. This contrary belief need not be a reasonable belief. To convict the defendant if he has an unreasonable belief about the existence of a factual element would be to convict him on a showing of mere negligence. Once we adopt an elements approach to the definition of offenses, to insist on privileging intentional-

109. *Id.* § 1.13(10). It is unclear what the drafters intended by this, or whether anything of significance can be inferred from it. The Code was written well before justifications and excuses were systematically distinguished in American law and legal theory.

110. *See* H.L.A. HART, THE CONCEPT OF LAW (1975) (condemning arguments that rely on definitions of words instead of analysis of concepts).

111. The other default rule is section 2.02(4), which provides that if a fault element is mentioned in the offense definition in a way that does not distinguish between factual elements, the fault element applies to all elements. MODEL PENAL CODE § 2.02(4). The principal difficulty is that section 2.02(4) reflects a "per offense" approach to criminal fault, instead of the novel and innovative "per element" approach reflected in section 2.02(1). *See* Robinson & Grall, *supra* note 40, at 715.

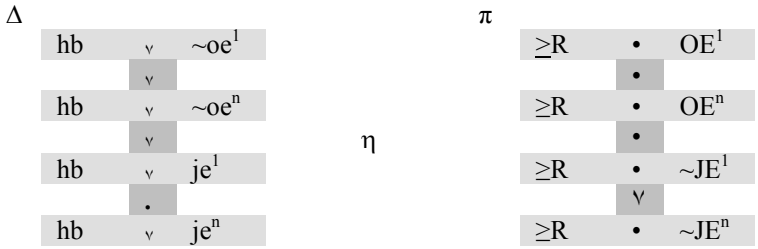
112. *See* MODEL PENAL CODE § 2.02(3) (providing that criminal fault is established for a given element "if a person acts purposely, knowingly or recklessly with respect thereto").

113. *See id.* § 2.04 cmt. 1 ("Thus ignorance or mistake is a defense when it negatives the existence of a state of mind that is essential to the commission of an offense . . . .").

states fault criteria in this way necessarily precludes a rule that a mistake must be reasonable to acquit.[114]

We know that the Code treats the elements of justification defenses as it treats the elements of offenses with respect to fault, because it treats mistakes about justification defenses as it treats mistakes about offenses. The Code's justification defenses require only a "belief," not a "reasonable belief," about justification elements to acquit.[115] This carries the requirements of 2.02(3) from the definition and proof of offenses into the definition and proof of justifications. If only an honest belief is required to support a justification defense, the prosecution cannot prevail unless it can show that the accused acted with some intentional state of mind inconsistent with such a belief. It will not suffice merely to show that the beliefs he did have were unreasonable. In other words, the prosecution must show that the accused was purposeful, or knowing, or advertently reckless with respect to his not being justified. Proof of negligence will not suffice. This is, of course, the move made in English law in *Beckford*.

The continuity view of criminal wrongdoing simply spells out this part of the Model Penal Code on the horizontal axes. Again, for reference purposes, here is the continuity view:



It is difficult to believe that the Code drafters imported 2.02(3), the default rule on fault elements, into the justification defenses without also importing the provision that makes it necessary: the requirement of 2.02(1) that the proof of each material element must be matched with the proof of an accompanying fault element. In order to prevail against a defendant's prima facie case on a justification defense, then, the prosecution must do more than show that the justifying circumstances did not obtain or that the defendant did not have a belief that they did obtain. The prosecution must show at least one such pair—fact element and intentional-states fault element—just as it must show paired fact and fault elements in the proof of the offense.

114. *Id*. ("There is no justification, however, for requiring that ignorance or mistake be reasonable if the crime or the element of the crime involved requires acting purposely or knowingly for its commission.").

115. *See, e.g.*, *id.* § 3.04(1) ("[T]he use of force upon or toward another person is justifiable when the actor believes that such force is immediately necessary for the purpose of protecting himself against the use of unlawful force by such other person on the present occasion.").

One objection to this interpretation of the Code is that, in the official commentary to section 3.04, which defines self-defense, the drafters say expressly that this is not how the Code's justification provisions work.

> The basic formulation of Subsection (1) requires that the actor believe that the circumstances create the necessity for using some protective force and that the force that he employs does not exceed what he believes to be essential to relieve his peril. Prevailing rules respecting self-defense, both common law and statutory, similarly demanded belief in the necessity of defensive action.[116]

A little further on, the commentary to section 3.04 states, "There is no case for acknowledging justification for the use of force against the person of another in the absence of belief by the actor in its necessity for the protective purpose he attempts to serve."[117] These passages suggest that a defendant must meet this evidentiary burden: hb • je. This is contrary to the continuity view, under which the defendant can prevail on a showing of hb $\vee$ je.

The problem is that the same commentary says exactly the opposite thing on the same or facing pages. The passage just quoted continues on:

> Equally clearly, a rule of liability dependent upon the actual facts, as distinguished from the facts as the actor believes them to be, is indefensible except to the extent, previously noted, that a prosecution for recklessness or negligence is justified based on the manner of the actor's perception. Otherwise, this is liability without culpability, which properly has no place in the penal law.[118]

If we are to avoid liability without culpability with respect to justifying circumstances, we must require the prosecution to prove fault in connection with non-justification, or $\geq$R • ~JE. The commentary reaffirms this commitment a few lines further down: "Legislative revisions since the adoption of the Model Code have, like the Code itself, rejected any strict liability approach to justifications for the use of force."[119] But, again, to avoid strict liability in the proof of justification defenses, we must require the prosecution to prove fault just as it is proved for offenses: $\geq$R • ~JE. From this, it appears that the Code adopts the continuity view: hb $\vee$ je η $\geq$R • ~JE.

The truth is that the Code adopts neither the *Dadson* view nor the continuity view. It adopts, instead, a third view that seems to resolve the central dilemma of mistakes about justifications—the choice between convicting the mistakenly unjustified actor and acquitting the mistakenly justified actor. Whereas the *Dadson* view treats beliefs about justifying circumstances as necessary for acquittal and the continuity view treats such beliefs as sufficient for acquittal, the Code adopts the position that a

116. *Id.* § 3.04 cmt. 2(a).
117. *Id.* § 3.04 cmt. 2(b).
118. *Id*.
119. *Id.*

belief in justifying circumstances is both necessary and sufficient for acquittal. On an elements table, this subjective view of the justification defenses looks like this:

Δ						π			
hb	∨	~oe[1]				≥R	•	OE[1]	
	∨						•		
hb	∨	~oe[n]				≥R	•	OE[n]	
	∨			η			•		
hb	∨	~~je~~[‡]				≥R	•	~~~JE~~[‡]	
	•						∨		
hb	∨	~~je~~[¤]				≥R	•	~~~JE~~[¤]	

On this view, the runner will be acquitted because she has a belief in justifying circumstances that is sufficient for an acquittal. The rogue detective will be convicted because he cannot show a necessary belief in justifying circumstances. Justifying circumstances as such are irrelevant to the case, except as the content of the defendant's beliefs.

For a number of reasons, however, the subjective view is not a viable solution to the central dilemma of mistakes about justification. In fact, the subjective view turns the dilemma into a trilemma—which the Code then mitigates but does not resolve. It is worthwhile to understand how this happens, in order to see that the Code drafters adopted the subjective view of justification defenses in part because they embraced the continuity view.

The sentences I quoted above that appear to support the continuity view are found in a discussion of a rule of prior law that the drafters were especially keen to replace. According to the law of New York, among other jurisdictions, an actor who acts unreasonably with respect to a justification loses the defense altogether—potentially resulting in a conviction for a crime done intentionally, which might carry a long sentence.[120] The Code drafters abolished this rule to avoid that result. The commentary's explanation for this move is a clear statement of the continuity view: "To convict for a belief arrived at on an unreasonable ground is to convict for negligence. Where the crime otherwise requires greater culpability for a conviction, it is neither fair nor logical to convict when there is only negligence as to the circumstances that would establish a justification."[121] If negligence is all that can be shown regarding a justification, conviction of an offense premised on purpose, knowledge, or recklessness ought to be precluded—in spite of the fact that the defendant actually did act with purpose, knowledge, or recklessness with respect to the elements of the offense. As we will see, this easy movement between offense and justification defenses—precluding the operation of one based on the proof regarding another—is a constant feature of the

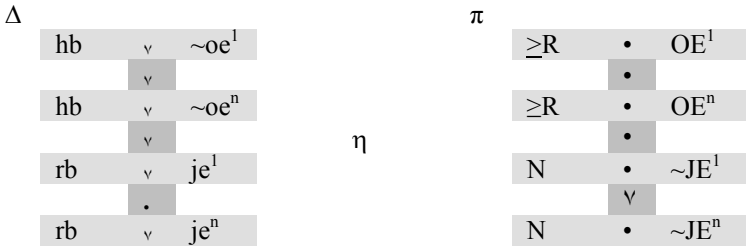
120. *Id.* § 3.02 cmt. 4.
121. *Id.* § 3.09 cmt. 2.

law on fault in homicide, beginning with the protean fault criterion of malice.

The Code drafters' rejection of the *Dadson* view and their commitment to a continuity view is equally clear in their statement of the general principles that they took to guide their drafting of the justification defenses. They write,

> A second principle is that the standards of culpability governing the absence of facts to justify the use of force should parallel the standards governing other material elements of an offense. Thus, the actor who negligently believes he must kill a person in self-defense should be treated like an actor who inflicts death by negligently discharging his gun. . . . Although some jurisdictions have followed the Model Code in correlating general standards of culpability to standards for justification, some others have not.[122]

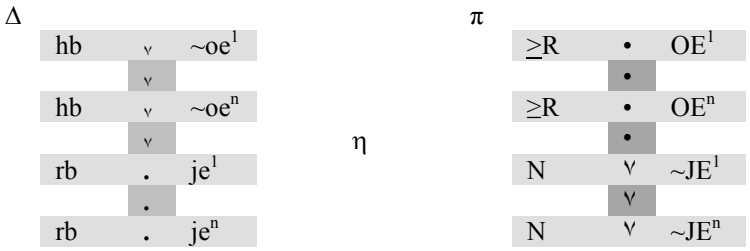
An actor who is accused of negligent homicide can defend himself by proving either that he did not act negligently *or* that he did not inflict death, because the prosecution is required to show that he did act negligently *and* that he inflicted death. If the proof of the justification defenses should parallel the proof of the elements of the offenses, the defendant can defend himself by proving either that he did not act negligently with regard to justifying circumstances *or* that justifying circumstances were present. This means that the prosecution is required to show that he did act negligently *and* that justifying circumstances were absent. In our terms:



This is the continuity view.

In contrast, the New York rule dictates that a belief in justifying circumstances is necessary but not sufficient to earn the defendant an acquittal. To say that negligence regarding justification will deprive the defendant of a justification defense and so serve to convict him is to say the prosecution can prevail on this showing: N ∨ ~JE. This implies that the defendant must show this to prevail: rb • je. In other words, the evidentiary burdens look like this:

122. *Id.* § 3.04 cmt. 1.



Of course, we have seen this before. In terms of elements, the New York rule reflects the *Dadson* view.

Read literally, however, the Code rejects the continuity view. Instead of treating a belief in justifying circumstances as sufficient to acquit, the Code makes such a belief a necessary condition as well.[123] But it is important to see that this is not a compromise with the *Dadson* view, under which a belief in justifying circumstances is necessary. The Code unambiguously rejects the *Dadson* view when it rejects the New York rule on mistakes in justification. The Code's rule that a belief in justifying circumstances is necessary also cannot be put down to its general commitment to a subjective or intentional-states conception of criminal fault. This accounts only for its treating belief as sufficient. The necessity of belief in justifying circumstances under the Code reflects an even deeper commitment to the idea that criminal desert is a matter of intentions. It is not enough that a justified act is not blameworthy; it must be praiseworthy as well. No praise is due to the accidental hero, such as the drunk on a yacht who throws a life preserver overboard as a joke, thus saving the drowning child who catches it. Likewise, no justification is earned by the unknowingly justified actor.[124]

The Model Penal Code's subjective view is problematic on its own terms, and in any event does not resolve the central dilemma presented by the two kinds of mistake about justification. Taken on its own terms, the defects in the Code's subjective view of the justification defenses run deep. Suppose a battered woman is under active attack from her abuser, and she assaults him in response. On the subjective view, her belief that she was under attack justifies the use of force. The fact that what she believes to be true *is* true is pleasantly reassuring, but entirely irrelevant. In this light, the "belief only" view of the Code's justification provisions seems wrong on its face. It says that justifying circumstances afford no justification, even in part. One need not accept the continuity view to see that this is so radically different from the relationship of fact to fault in offenses as to be utterly implausible. If we understand justification defenses as representing, in any fashion, the absence of criminal wrong-

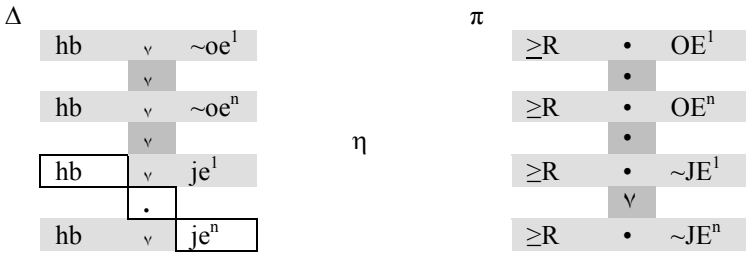
123. *See* Dillof, *supra* note 3, at 1555.

124. *Cf. id.* at 1570–71 (attributing this view to "desert-based theories of criminal liability," though not to the Model Penal Code as such).

doing, then non-justification as a matter of beliefs alone looks uncomfortably like punishment for thoughts.

Suppose that the battered woman uses deadly force in response to the attack and that she does this, not because she believes the use of deadly force is immediately necessary, but because she already has a gun in her hand and firing it is the most effective way to repel the attack. If her attacker had a gun that she simply did not see, the use of deadly force was, in fact, immediately necessary. On the "belief only" view, she cannot claim self-defense in this situation because the required belief is not only sufficient but also necessary. But she ought to be allowed to argue self-defense. Her *belief* about the immediate necessity of the use of force and the *fact* that deadly force is required can and ought to be *combined* to afford her this defense.

This table shows her argument and its viability on the continuity view:



All justifying circumstances are present—she did in fact face a deadly threat—and all justification elements can be proved. Even if these justifying facts could not be proved, the prosecution could not prove fault with regard to any absent justifying circumstance—except for one. And it consists of a thought. The defendant thought the threat was less grave than it turned out to be. In every other respect, she has behaved as she ought to have, and she is genuinely innocent even though she has killed intentionally. To reject her defense on the sole ground of her belief in a non-deadly threat, when she is innocent in every other respect, is not punishment in the absence of any act at all—a true case of punishment for thoughts—but it seems only slightly less problematic.

This argument against the subjective view is open to debate, of course. What is not open to debate, it seems to me, is that the Model Penal Code does not resolve the central dilemma presented by cases such as the runner's and the rogue detective's. Its subjective view compounds the problem, creating a trilemma. The Code's subjective view of mistakes in justification creates a problem of the idiosyncratic believer. If the concert master kills the conductor because he honestly believes the conductor is about to stab him to death with his baton (because he played the A flat), the concert master is entitled to an acquittal. If belief in justifying circumstances is both necessary and sufficient, the only other variable—the justifying circumstances themselves—can be neither suffi-

cient nor necessary. And by depriving the underlying facts of any role in the analysis, the subjective view cuts the justification defenses loose from any anchor in the real world.

To mitigate the problem of the idiosyncratic believer, the Code drafters offered a rule under which negligently maintaining a belief in justifying circumstances deprives the defendant of a mistake defense for a crime of negligence; and recklessly maintaining a belief in justifying circumstances deprives one of a mistake defense for crimes of recklessness.[125] This rule, however, has three weaknesses. First, it does not cure the idiosyncratic believer problem; it merely mitigates it. The concert master will in any event be acquitted of murder for his purposeful killing. Second, the rule functions rather badly. The concert master could be shown to have recklessly maintained his erroneous belief only by some implausible tale such as this one: "I've often thought I might have to kill the conductor before he could stab me to death for playing the A flat, but I thought my position in the orchestra was worth that risk." Finally, by offering a rule meant to capture the idiosyncratic believer, the Code's drafters merely affirmed the principal shortcoming of their approach to the problem and confirmed the intractability of the basic dilemma. Unless we are satisfied with the Code's mitigation of its implausible subjective view of justification defenses, we must convict the runner in order to convict the rogue detective; and we must acquit the rogue detective in order to acquit the runner.

Because the Code's subjective view does not resolve the basic dilemma this Article addresses, I will have nothing more to say on it. Instead, I will pursue the continuity view—a view the Code drafters adhered to, even though a stronger concern on their agenda overrode it.

IV. THE CASE FOR THE CONTINUITY VIEW

The central claim of the continuity view of the justification defenses is that a person who makes a mistake about the justifiability of his actions is genuinely justified. Conventional theory has it that the mistaken actor is "merely excused." I made my case against this kind of excuse—a defense purportedly distinct from offense, justification, and nonresponsibility, but otherwise unanalyzed—in Part II of this Article. The question now is whether I can show that the mistaken actor is genuinely justified. The case that I will offer for the continuity view of justification defenses is historical and doctrinal where the horizontal axes are concerned, conceptual where the vertical axis is concerned, and formal where second-best solutions are concerned.

In Part IV.A, I will argue that offense and justification defense are continuous along the horizontal axes connecting fault and factual ele-

125. MODEL PENAL CODE § 3.09(2).

ments. The Model Penal Code reflects the continuity view—which is hardly surprising given that it is largely drawn from the Code—but so does non-Code law on justification defenses. Non-Code law in this area turns on the idea of malice, and the history and interpretation of this fault criterion supports the continuity view. Malice is hardly the malignant legal norm it is often taken to be.[126] Malice is—as it plainly has been throughout the history of criminal law—a paradigmatic fault criterion. More important, malice runs through offense and justification defense without differentiation. Even as both Code and non-Code law reduce malice to more specific fault criteria, they continue to display this pattern.

The conceptual case for continuity offered in Part IV.B deals with the flip in the vertical connector between paired elements, from "∨" to "•" on the defense side, and from "•" to "∨" on the prosecution side, as the proof moves from the offense to the justification defense. The objection is that, because of this flip, the analysis could just as well be called "the discontinuity view" of criminal wrongdoing, in a way that casts doubt on the idea that a mistake about a justification is a case of genuine justification. This argument is bolstered by the fact that a justification defense is a second rule that cancels the operation of the first rule, the prohibition. Furthermore, this second, cancelling rule has its own purposes or ends that are distinct from those of the first rule. However, the logic of exceptions to prohibitions, even if they are construed as cancelling permissions, supports the continuity view. An answer to the second objection—that offenses and justification defenses have different or competing ends—takes us into Aristotelian punishment theory, introducing one of its most important concepts: the law's specification of competing ends.

The principal alternative to the continuity view is the *Dadson* view, which has the advantage of handling the case of the mistakenly justified actor handily. This advantage shows up in formal terms in a flip in the connectors on the horizontal axes that matches the flip in connectors on the vertical axis—a formal difference that supports the doctrinal case and the normative case for discontinuity between offenses and justification defenses offered by Simester and Gardner, respectively. However, the *Dadson* view reaches a counterintuitive result regarding the mistakenly unjustified actor. The *Dadson* view and the continuity view each have a second-best solution to deal with their respective troublesome cases.

126. *See, e.g.*, Eric A. Johnson, *The Crime that Wasn't There: Wyoming's Elusive Second Degree Murder Statute*, 7 WYO. L. REV. 1, 2 (2007) ("Further, the current definition of malice—which requires the state merely to prove *either* that the defendant acted 'without legal justification or excuse' *or* that the defendant acted with 'hatred, ill will, or hostility'—is a throwback to the unhappy days when judges used the word 'malice' in 'the old vague sense of "wickedness in general."'"); Jeremy M. Miller, *Mens Rea Quagmire: The Conscience or Consciousness of the Criminal Law*, 29 W. ST. U. L. REV. 21 41–42 (2001) (arguing, against malice, that "defining mens rea as evil mind, and including it as an element of the crime *was and is* a *primitive* mode of crime definition").

The difference, I will argue in Part IV.C, is that the *Dadson* view has stronger counterintuitive implications than is generally recognized—none of which is avoided or mitigated by its second-best solution.

A. The Continuity of the Horizontal Axes

The drafters of the Model Penal Code were gifted scholars, but it is hard to credit them with uncovering the true structure of criminal wrongdoing, and even harder to take Code doctrine as definitive on theory. The continuity view of criminal wrongdoing is evident in the Code's fault and justification provisions because the continuity view is a description of these provisions. They hardly count as a reason to adopt it. On the other hand, it is easy to expand this vicious circularity into virtuous coherence. Taken as a restatement of criminal law, the Code's fault and justification provisions capture well the common law of crime and early codifications. Just as the Code treats mistakes about justification in terms of fault and as continuous with the proof of fault in offenses, so did prior law. If the continuity view accords with this body of law, we are on firm ground in taking that accord as sufficient reason to adopt it. The Model Penal Code might give an idiosyncratic and inaccurate account of fault, justification, and mistake, and this might infect the continuity view. But one can hardly say the same of the continuity view's accord with the common law of crime and early codifications. They are all the data there is to be explained.

In the law of homicide, malice is the mother of all fault criteria. Courts have used and still use malice to describe both intentional-states and non-intentional-states fault in homicide: both the intent to kill and the depraved heart of the murderer constitute malice. Malice was and is used to distinguish murder from manslaughter: murder includes it and manslaughter does not.[127] Except, of course, when it does. An intentional killing is not murder, but manslaughter instead, if some other circumstance demonstrates the absence of malice—never mind, apparently, that an intent to kill constitutes malice.[128] Illogical as it might seem, however, the protean nature of malice reflects the continuity of offense and defense with respect to criminal fault's intentional and non-intentional criteria.

Fourteenth century cases on self-defense in homicide treated it as an excuse negating malice.

127. Perkins dates this distinction to the fifteenth century: "A series of statutes, during the period from 1496 to 1547, excluded from benefit of clergy certain of the more serious forms of felonious homicide, referring to them as murder committed with malice aforethought (malice prepensed). . . . The word 'manslaughter' was adopted to describe this lesser grade of felonious homicide." Perkins, *supra* note 15, at 543–44.

128. *See id.* at 538–39 ("'Malice' . . . is frequently used in the law to mean 'the intentional doing of an unlawful act.' But many such acts are not done with malice aforethought.").

> The phrase "malice aforethought," however, had made its appearance in the law long before this legislation had put certain types of felonious homicide beyond the protection of benefit of clergy. But strangely enough, its original meaning was altogether different from that with which it was used in these statutes. For example, in 1329, the jurors, in bringing in a verdict of conviction of excusable homicide, included in their finding—in order to entitle the defendant to a pardon—that he had killed the deceased "in self-defense, and not by felony or of malice aforethought."[129]

Notice two things about this law. First, it employs a meaningful sense of the term "excuse." An excused crime was one for which one might receive a pardon, thus avoiding corporal punishment, but conviction for which would, nevertheless, result in a forfeiture of property.[130] This is why the defendant was said to be convicted of excusable homicide. This excuse of self-defense has no relation to the unanalyzed category of excuse now used to describe one who is mistaken about self-defense. Second, and more important, notice that this law treated self-defense itself as negating malice. That is, the defense taken as a whole—and not, for example, mistake about self-defense—negated the non-intentional fault element of the offense, malice. Self-defense has been conceptualized, from the beginning, as part of a structure of criminal wrongdoing in which there is continuity between offense and defense on the question of fault.

The doctrine of imperfect self-defense perpetuates this view of criminal wrongdoing, in that it treats an unreasonable mistake about justification as being inconsistent with the malice required for murder, resulting in a conviction for manslaughter.

> Logically, a defendant who commits a homicide while honestly, though unreasonably, believing that he is threatened with death or serious bodily harm, does not act with malice. Absent malice he cannot be convicted of murder. Nevertheless, because the killing was committed without justification or excuse, the defendant is not entitled to full exoneration. Therefore, as we see it, when evidence is presented showing the defendant's subjective belief that the use of force was necessary to prevent imminent death or serious bodily harm, the defendant is entitled to a proper instruction on imperfect self-defense.[131]

129. *Id*. at 544–45.

130. *Id*. at 540 ("Homicides which entitled the slayer to a pardon, even though they resulted in conviction and forfeiture of goods, were spoken of as having been committed 'without felony'—they were pardonable or 'excusable.'").

131. State v. Faulkner, 483 A.2d 759, 768–69 (Md. 1984); *see also* People v. Randle, 111 P.3d 987, 991–92 (Cal. 2005) ("One acting in imperfect self-defense also actually believes he must defend himself from imminent danger of death or great bodily injury; however, his belief is unreasonable. Imperfect self-defense mitigates, rather than justifies, homicide; it does so by negating the element of malice."); State v. Thomas, 114 S.E. 834, 847 (N.C. 1922) ("But if the slayer acts from an honest belief that it is necessary to protect himself, and not from malice or revenge, even though he formed such conclusion

Notice the easy movement between defense and offense in this argument. An unreasonable mistake about justification is nevertheless a mistake about justification, not about the offense. And yet the mistake affects the offense, negating one of its elements. This element is malice, a species of fault that apparently runs continuously through the offense and the justification defense, so that a mistake concerning one precludes conviction for the other.

The reasons for this continuity need to be clarified, but another problem commands our attention first. Imperfect self-defense does not result in no crime, as the negation of malice in murder implies. It results in a lesser crime—manslaughter instead of murder. The problem is that this seems to be accomplished by main force, with courts and legislatures relying on the latitude afforded by "malice." Given that it means anything we wish it to mean, its negation can have any implications we like.

Consider another ground of imperfect self-defense. Some jurisdictions have recognized excessive use of force in self-defense as supporting a manslaughter conviction instead of either murder liability or full acquittal.[132] The stated ground for this alternative is the absence of malice where excessive force is used.

> The circumstances which attended the killing may, however, be shown to rebut the presumption of malice. This may be done by a showing that the homicide was committed in self-defence and was therefore excusable, or by a showing of circumstances which although they would not excuse or justify the act would mitigate the crime from murder to manslaughter. It is the last proposition which primarily concerns us.[133]
>
> . . . The jury, however, should have been instructed further, upon the assumption that the deceased was the original assailant, that if the use of the knife by the defendant as a means of averting harm to himself was unreasonable and clearly excessive in light of the existing circumstances, they could conclude that the defendant himself became the attacker and, since death resulted from his use of excessive force, he would be guilty of manslaughter.[134]

As with mistake of fact with respect to a justification, manslaughter premised on excessive use of force is analyzed as neither justification nor excuse, confirming that the malice negated is that of the offense of murder. However, in spite of this similarity and the similarity in results under these doctrines of imperfect self-defense, the analogy from excessive force to mistake is flawed. It suggests that, like mistake, excessive force is inconsistent with malice. But the opposite seems true. Excessive force

hastily and without due care, and when the facts did not justify it, still, under such a case, although such belief on his part will not fully justify him, it may go in mitigation of the crime, and reduce the homicide from murder to manslaughter." (emphasis omitted)).

132. *See, e.g.*, Commonwealth v. Boucher, 532 N.E.2d 37, 40 (Mass. 1989).

133. Commonwealth v. Kendrick, 218 N.E.2d 408, 413 (Mass. 1966).

134. *Id*. at 414.

has no exonerating implications; it seems to confirm, not to negate, malice.

At best, malice and its negation are extraordinarily capacious and malleable terms. The rationale for excessive force manslaughter seems to be that self-defense as a whole negates the malice of murder, resulting in no crime. Excessive force then raises liability from nothing to manslaughter, as the court says in the passage quoted above. A principled rationale for these moves is hard to come by. In fact, the rationale seems incoherent. If proportionate use of force is a necessary element of self-defense, then excessive force simply precludes a finding of self-defense and leads to a conviction for murder, not acquittal. Even if we assume that self-defense by means of excessive force acquits the defendant of murder, it is difficult to see how the use of excessive force in an intentional killing could premise manslaughter.

One likely rationale, such as it is, is the same as that for imperfect self-defense premised on mistake. Manslaughter is a second-best solution to a choice between unpalatable alternatives: conviction for murder or full acquittal. This course seems unprincipled in both doctrines. Indeed, we have an unprincipled solution to a non-problem. One who uses excessive force in self-defense or who is mistaken about the necessity of self-defense should simply be convicted of murder. With imperfect self-defense based on the use of force, we are even worse off. Excessive force, unlike mistake, seems to confirm malice, not to negate it.

At this point, the law of malice seems to offer support for the continuity view that it would be better off without. The continuity of malice in murder and its negation leading to manslaughter liability is unprincipled and illogical. It seems stipulative, making the thesis that criminal wrongdoing is continuous between offense and defense less a theoretical description than a bit of sociology.

To begin to rehabilitate the continuity thesis, we should reconsider the analysis of imperfect self-defense premised on the use of excessive force. I said that excessive force confirms malice instead of negating it; but this is not entirely true. One way to look at the use of excessive force is as an indication of a momentary loss of rational control—a frenzy brought on through no fault of the actor. It might confirm malice in most contexts, but not so where the actor has been threatened with the unlawful use of force by another person. In this context, a momentary loss of control is understandable, and to cap criminal liability at manslaughter for what would otherwise be murder seems fair.

This rationale recalls provocation, of course, and this is my point. Provocation manslaughter is often described as resulting from the negation of the malice required for murder. Federal criminal law frames it this way:

> In this jurisdiction, a homicide constitutes voluntary manslaughter where the perpetrator kills with a state of mind which, but for the

> presence of legally recognized mitigating circumstances, would render the killing murder. This definition of voluntary manslaughter reflects the traditional common law view and the prevailing national norm, as indicated by the formulations in numerous criminal law treatises. *See, e.g.*, 2 C. TORCIA, WHARTON'S CRIMINAL LAW § 153 (14th ed. 1979) (for voluntary manslaughter liability to attach, the "defendant may act with the *intent to kill or with any mental state which amounts to 'malice'*; the malice is negated by the provocation and the offense is mitigated from murder to voluntary manslaughter" (emphasis added)).[135]

From this perspective, provocation forms part of a triumvirate of sorts—together with mistake regarding justifying circumstances and the use of excessive force—of malice-negation defenses. Here, for example, is the Supreme Court of North Carolina grouping provocation together with mistake regarding justifying circumstances, in 1922:

> Certainly fright or terror will not excuse the unnecessary taking of human life when there is no reasonable ground for apprehending death or enormous bodily harm, but, in connection with other circumstances, it may serve to repel the inference of malice arising from the intentional killing with a deadly weapon and to mitigate or reduce the homicide from murder in the second degree [to] manslaughter. . . .[136]
>
> . . . It is not always necessary to show that the killing was done in the heat of passion to reduce the crime to manslaughter, for where the killing is done because the slayer believes that he is in great danger, but the facts do not warrant such belief, it may be murder or manslaughter according to the circumstances, even though there be no passion.[137]

And here is the Court of Appeals for the District of Columbia equating provocation with the use of excessive force, in 1984:

> This type of situation will arise "where the killer has been provoked or is acting in the heat of passion." It also may occur when excessive force is used in self-defense or in defense of another and "[a] killing [is] committed in the mistaken belief that one may be in mortal danger." In each of these situations—where the "purpose to kill is . . . dampened" so as to "mitigate malice,"—a defendant is guilty of voluntary manslaughter, not murder.[138]

One might expect cases to arise that would support all three theories on the facts, and they do.[139]

135. Comber v. United States, 584 A.2d 26, 43 (D.C. 1990) (footnotes and some citations omitted).

136. *Thomas*, 114 S.E. at 836.

137. *Id*. at 837.

138. Logan v. United States, 483 A.2d 664, 671 (D.C. 1984) (citation omitted).

139. *See, e.g.*, State v. Clark, 77 P. 287 (Kan. 1904); Johnson v. State, 2 S.W. 609 (Tex. App. 1886).

Occasionally a court will draw the conceptual connections between the three doctrines. In *Commonwealth v. Boucher*,[140] the defendant and his friends blocked the path of the victim's car. Once he got around them, the victim got out of his car, ran toward the defendant, and attempted a "karate kick" to the defendant's head. In the ensuing fight, the defendant stabbed the victim through the heart. The trial court instructed the jury on murder, manslaughter, provocation (premised on mutual combat), and self-defense, but failed to explain that a finding of provocation negated malice and precluded a conviction on murder. In explaining the errors in these instructions, the Supreme Judicial Court of Massachusetts treated provocation as implicit in mistake regarding self-defense.

> Inherent in the defendant's argument that he was not guilty because he had acted in self-defense was his reliance on provocation to mitigate the charge of murder to manslaughter, if the self-defense claim were to fail. Thus the judge's omission created the significant possibility that the defendant was erroneously convicted of murder in the first degree instead of manslaughter.[141]

The appellate court also found error in the failure to instruct on manslaughter premised on the use of excessive force in self-defense. It described this defense, too, as conceptually related to provocation by the idea that malice is negated in such cases.

> The judge's two omissions in dealing with provocation in his charge are conceptually related. The provocation that justifies reasonable action in self-defense also negates a finding of malice in any killing that results from the use of excessive force. It is just one example of the general proposition that reasonable provocation negates malice in any unjustified killing, warranting a verdict of manslaughter rather than murder. In neither instance did the judge adequately explain the mitigating influence of provocation.[142]

Boucher, however, provides no account of this conceptual relationship beyond noting it.

The conceptual relationship between imperfect self-defense and provocation is best described in terms of the continuity view of criminal wrongdoing. In Massachusetts, the "defense" of provocation is found in the definition of manslaughter,[143] just as the "defense" of extreme mental

140. 532 N.E.2d 37, 40 (Mass. 1989).

141. *Id*. at 39–40.

142. *Id*. at 40; *see also* Commonwealth v. Vinton, 733 N.E.2d 55, 64 (Mass. 2000) (approving an instruction stating, "A person in fear of imminent injury or death who uses excessive force to defend himself, would be acting under a form of provocation, and that would reduce the crime from murder to manslaughter.").

143. Massachusetts has no statutory definition of manslaughter but provocation is consistently included in the common law definition of the offense. *See* 32 MASS. PRAC., CRIMINAL LAW § 202 (2007) (citing Commonwealth v. Soaris, 175 N.E. 491, 494 (1931)) (defining voluntary manslaughter as "a killing from a sudden transport of passion or heat of blood, upon a reasonable provocation and without malice, or upon sudden combat").

or emotional disturbance is found in the Model Penal Code's definition of manslaughter.[144] In this respect, Massachusetts law accords with other non-Code jurisdictions that define provocation by reference to murder and the negation of malice.[145] In contrast, the doctrine of imperfect self-defense, whether premised on mistake or excessive force, is not contained in any offense definition. Both of these mitigations are defined by reference to self-defense, of course. But the rationale of imperfect self-defense is the same as the rationale of provocation: the negation of an offense element, malice, in murder.

Because the three doctrines share a common rationale that turns on criminal fault, the homicide offenses and justification defenses such as self-defense are continuous with respect to fault. That is, the operation of fault with respect to offense elements is identical to its operation with respect to elements of a non-justification, regarding both mechanism and result. A successful showing of a non-intentional fault element in the offense—adequate provocation—negates the fault element of murder: malice. Likewise, a successful showing of a non-intentional fault element for non-self-defense—e.g., negligence regarding non-necessity due to mistake—negates the fault element of murder: malice. And again, a successful showing of a non-intentional fault element for non-self-defense—negligence regarding disproportionality due to the use of excessive force—negates the fault element of murder: malice. Each of these showings of non-intentional fault results in manslaughter liability, regardless of whether it operates in an offense definition or in a justification defense. Offense and justification defense work in strictly analogous fashion in these homicide doctrines, leading to the same result because offense and justification defense are continuous with respect to non-intentional fault.

The Model Penal Code follows a similar pattern. The differences between the non-Code and Code provisions relate to the Code's strong preference for intentional-states fault criteria, in contrast to non-Code law's use of malice. What matters for our purposes, however, is that these doctrines all concern fault, both intentional and non-intentional. Extreme mental or emotional disturbance is a non-intentional fault criterion that appears in the definition of manslaughter. As I explained above, the non-intentional elements of EMED supplement the intentional-states definition of murder. Some murders are purposeful or knowing killings that are no worse, morally, than reckless killings. The alternative, EMED, definition of manslaughter captures these otherwise overincluded murders. Similarly, the Code's version of imperfect self-defense consists of non-intentional fault criteria that appear in qualifications to the Code's definition of justification defenses. An honest belief is sufficient to constitute an exonerating mistake of fact under the Code's

144. MODEL PENAL CODE § 210.3(1)(b) (1962).

145. *See, e.g.*, Morgan v. United States, 363 A.2d 999, 1002 (D.C. 1976).

justification defense definitions, and this results in underinclusion. An idiosyncratic believer will be acquitted where a requirement of reasonable belief would convict. Sections 3.02(2) and 3.09(2) ensure that a dangerous or malicious actor is convicted of manslaughter or negligent homicide—usually known as involuntary manslaughter—if his belief is unreasonable.

Whether it appears in an offense definition or a justification defense, each of these qualifications backs up, so to speak, an intentional-states fault doctrine that is over- or underinclusive relative to our moral judgments on the cases. Together, these intentional-states fault criteria and their qualifying non-intentional fault criteria operate just as protean malice does. They provide a supple law of criminal fault that serves to match legal judgments to moral judgments on wrongdoing. As in prior law, these showings of non-intentional fault result in manslaughter liability under a single rationale that runs continuously from offense definitions to justification defenses.[146]

Given the continuity of offenses and justification defenses with respect to fault, it would be remarkable if we were to treat the relationship between fault elements and fact elements differently in the definition of offense and defense, respectively. If the prosecution must prove the conjunction of fault and fact in the proof of an offense, then the prosecution must prove the conjunction of fault and fact with regard to non-justification. If the defendant can rebut the prosecution's case on the offense by meeting a burden of production and raising a reasonable doubt about either a fault element or a fact element in the offense, then she can rebut the prosecution's case on non-justification by meeting these evidentiary burdens on either a fault element or a fact element in the justification defense.

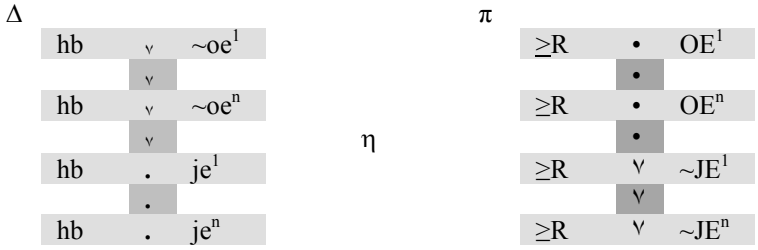
B. Continuity Along the Vertical Axis

Some courts and scholars say that a defendant such as the rogue detective is simply guilty of murder. This is the *Dadson* view of criminal wrongdoing and the justification defenses.[147] To be acquitted, the actor must prove (that is, make a prima facie case and raise a reasonable doubt as to) both the non-justification elements and also an intentional state of mind regarding each of these elements. He must, furthermore, do this for all elements of the justification defense.

146. Notice, furthermore, that the Model Penal Code replicates the non-Code law at the cost of some plausibility. Killings done under extreme mental or emotional disturbance or that are justified by an honest belief remain murders. Neither doctrine operates by means of the actual negation of the offense element of purpose or knowledge. They both seem merely stipulative as a result. What is noticeable about this, for our purposes, is that this problem arises with respect to both offense and defense, and that we have made the same choice, to accept this implausibility, in both offense and defense definitions.

147. *See* Christopher, *supra* note 5, at 229.

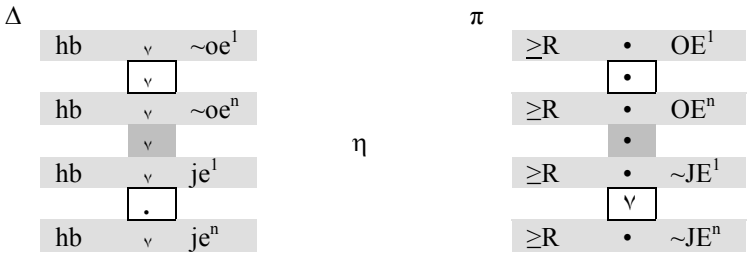
The *Dadson* view is this:



The detective ought to be convicted because the prosecution can meet its evidentiary burden of showing only that the defendant is at fault regarding the absence of justifying circumstances. Because he believes that there are no justifying circumstances, he is at least reckless, ≥R, as to non-justification, ~JE^n. Justifying circumstances are present, and so the prosecution cannot prove the factual elements ~JE^1 . . . ~JE^n, but this is immaterial. For the prosecution, the fault and fact elements regarding non-justification are horizontally and vertically disjunctive, and the proof of either one in any pair will suffice to convict.

The second part of my case for the continuity view consists of a contrast in the respective defenses of *Dadson* and continuity—an indirect argument against the *Dadson* view, but a strong one. (I will offer direct arguments against the *Dadson* view in Part IV.C.) *Dadson* and the continuity view agree that the connector on the vertical axis flips as the proof moves from offense to justification defense. But, because it suggests a fundamental difference between offense and defense, the flip poses a significant challenge to the continuity view—a challenge seemingly supported by two powerful analyses of justification defenses as exceptions to rules, advanced by John Gardner and Claire Finkelstein. The capacity to avoid a powerful objection is a measure of the strength of any theory. My analysis of the vertical axis reinforces the continuity thesis by showing that it accords with the logic of cancelling permissions, as Gardner describes justifications, and by providing an explanation for continuity in terms of the ends we pursue in the imposition of legal punishment, an explanation in terms of Finkelstein's analysis of exceptions. Because my analysis of the vertical axis supports the continuity view, it accords with my analysis of the horizontal axes. Supporters of the *Dadson* view have not produced a theory explaining the relationship of vertical to horizontal as they describe it. We have a right to expect such an explanation because it is no more obvious that both horizontal and vertical connectors ought to flip than it is that only one of them should.

On the same elements analysis that shows continuity on the horizontal axes, a difference between offense and defense is clearly indicated by the flip in connectors between paired elements on the vertical axis, from "∨" to "•" on the defense side, and from "•" to "∨" on the prosecution side. This feature is highlighted here:



As I noted at the end of Part III.A., this formal difference goes along with the normative differences between offenses and justifications described by John Gardner, and the doctrinal differences noted by Andrew Simester. So why should we call the result of an elements analysis "the continuity view," instead of the apparently more apt "discontinuity view"? Why not treat the vertical connectors as dispositive on the question of genuine justification, reject the idea that mistake in offense and defense have the same implications for guilt, and describe a case of mistake about a justification as anything but genuine justification—whether it be "excused" or something else?

The answer is that the flip in connectors along the vertical axis is epiphenomenal to the continuity view. This is by no means to say that the table does not show a real difference between offense and defense. The vertical axis reflects the facts that a justification, taken as an exception to the prohibition, is a second rule that cancels the first, and that this exception serves principles or ends that conflict with the prohibition. As I will argue here, however, neither one of these features undermines the continuity view. Even when it is construed as a second rule that cancels the first, an exception to a rule operates only within the boundaries of the rule it cancels. As for the conflict in principles or ends that differentiates the prohibition and its exception, an Aristotelian analysis of conflicting ends' reconciliation shows this reconciliation's occurring in a continuous fashion across offense and justification defense.

1. Exceptions, Cancelling Permissions, and Conflicting Ends

Almost inevitably, we think of justification defenses as exceptions to criminal prohibitions. An exception to a legal rule, however, is a tricky notion. The principal question is whether a rule with an exception is one rule or two. Presumably, the continuity view is better supported by the idea that an exception is part of a single rule, and the challenge to that view, stressing discontinuity, is supported by the two-rule position. But whether and why this is so is not altogether clear.

An exception might be seen as part of a single rule if its separation from the rule is a fortuity of language and not a matter of the rule's sub-

stance. Frederick Schauer argues that this is always the case.[148] But this is not to say that exceptions have no substantive rationale. Schauer describes exceptions to rules as overrides. If a legal rule could be set aside whenever its background justification were inapplicable, we would have no rule at all; we would have returned to guiding action on an ad hoc, all things considered basis.[149] We do set rules aside, but only according to rules that guarantee the rule is overridden in relatively rare and truly exceptional circumstances.[150]

John Gardner argues, to the contrary, that an exception to a legal rule is a second rule that cancels the first, and not simply a temporary override. Ordinarily, we act in favor of the balance of reasons. A rule is a second-order reason—a reason about reasons—in that it requires us to act not in favor of the balance of reasons, but to act, instead, according to the terms of the rule.[151] An override occurs when the ordinary balance of reasons strongly favors acting against the rule.[152] In other words, an override relies on first-order reasons. When the underlying balance of reasons returns to a semblance of normality, the rule again obliges us. This seems to be a perfectly good description of informal, moral justifications. But when we deal with legal rules and justifications, we do something different. A legal rule is a protected reason for action, in that it retains its normative force no matter how strongly the balance of first-order reasons compels us to act against the rule.[153] Only another legal rule can permit a departure from it. But if a second rule does this, then the first one is not merely overridden; it is cancelled.[154] Only the terms of the second rule or a third one, not a return to normality in the underlying balance of reasons, can reinstate it.[155]

In the defense of lesser evils, for example, the jury weighs the reasons for the defendant's actions, but this does not make the defense itself a matter of overriding reasons. Given the balance the jury arrives at and the validation or rejection of the balance drawn by the defendant, the law requires acquittal or conviction in no uncertain terms. If the decision is for an acquittal, the prohibition is cancelled and cannot be applied to the case. There is no further weighing of the underlying reasons for action—as there would be, for example, if the jury could skip over the question of acquittal or conviction and impose a penalty that reflected

148. Frederick Schauer, *Exceptions*, 58 U. CHI. L. REV. 871, 872 (1991).

149. *Id*. at 894–95.

150. *Id*. at 896–97.

151. John Gardner, *Justifications and Reasons*, *in* HARM AND CULPABILITY, *supra* note 6, at 103, 114; *see also* JOSEPH RAZ, THE AUTHORITY OF LAW 18, 21–33 (1979) (describing rules as protected reasons).

152. Gardner calls this a defeated rule. *See* Gardner, *supra* note 151, at 108–09.

153. *Id*. at 115.

154. *Id*. at 116–17.

155. *Cf. id.* at 117 ("In the case of a criminal act, all reasons are defeated apart from those permitted by law.").

the degree of the defendant's departure from the correct balance of reasons.

Claire Finkelstein makes an argument against Schauer that dovetails with Gardner's description of justifications as cancelling permissions. She argues that we do not impose exceptions where the background purpose is not served by the rule or would be better served by not imposing the rule.[156] Instead, we recognize an exception when we have a positive reason to do so: when a different background justification intrudes, making a choice or reconciliation of rules necessary.[157] An override as Schauer describes it is not an exception at all, and a justification cannot be an override if it is an exception. An exception involves two principles. In cases of self-defense, she argues, the prohibition on murder or assault serves to deter, to impose just deserts, and so on. But a justification defense serves a distinct and conflicting interest in fundamental political values such as autonomy.[158] The prohibition conflicts with liberty, and its enforcement conflicts with our rights to freedom from the restraint or death entailed in punishment. Finkelstein's point fits nicely with Gardner's. If a cancelling permission does not turn on the balance of reasons underlying the rule cancelled, then it might often be premised on a principle that competes with the principle behind the rule.

If a justification defense is a cancelling permission that responds to a different set of reasons than the prohibition does, it seems odd, at least, to say that justification defenses are continuous with offenses. On closer examination, however, it is not odd at all. Both Gardner's and Finkelstein's points support the continuity view. Let me take them in turn.

2. Cancelling Permissions and the Continuity of Justifications

We can reconcile Gardner's view of justifications as cancelling permissions with the continuity view of justification defenses if we think more carefully about how a cancelling permission works.

A rule is a protected reason. If a justification is a cancelling permission, and if what is cancelled is the protected status of the reason constituting the prohibition, then a justification defense seems to return us to governance by first-order reasons in the domain of the prohibition. In homicide, these first-order reasons include both those for and against killing another human being; in theft, both those for and against taking property, and so on. A cancelled prohibition no longer operates at all regarding these reasons. Of course, a justification defense operates to guide conduct as well as to cancel the prohibition, and so cases covered by the cancellation are not actually returned to the mix of first-order rea-

156. Claire O. Finkelstein, *When the Rule Swallows the Exception*, 19 QUINNIPIAC L. REV. 505, 514–15 (2000).

157. *Id*. at 515.

158. *Id*. at 531.

sons. But the fact that a justification defense governs the same first-order reasons behind the prohibitions on murder, theft, and so on, means that a justification defense prohibits by implication. The terms of such implied prohibitions are not as precise as the familiar prohibitions on murder and theft, but they are no less binding for that. They consist of the protected reasons of the justification defense, which operate relative to the same first-order reasons otherwise covered by express prohibitions.

This seems wrong. It is true that a justification defense operates relative to the same reasons that prohibitions do, and that to some extent justifications imply prohibitions—but this can be so only to a limited extent. It seems that, unless we suppose some general background prohibition, then any prohibition implied by a justification defense would not simply lack the detail of our express prohibitions; it would lack any force at all. Most conduct within that practical context would fall back into the mix of first-order reasons.

In fact, however, the implied prohibitions are likely to be well defined. If we protect reasons for some act, then, within some more or less well defined practical context, we protect the reasons against it as well. This picture fits with the ordinary operation of permissions. Suppose that I have never prohibited my daughter from using my car. If I suddenly announce that I am giving her permission to use my car on Tuesdays (when she has late soccer practice), she will ask me why she needs permission. If I say she needs permission because she is prohibited otherwise, she will say, "Well, obviously, but why?" She will ask, in other words, for the first-order reasons behind the prohibition. That there is a prohibition will be taken as given once I announce the permission. She might try to argue that she can also drive the car on Wednesdays for various reasons. But she would have to address not only these first-order reasons for her driving on Wednesdays, but also the first-order reasons against her driving on Wednesdays. If I simply assert the prohibition at this point, and refuse to debate the first-order reasons for her driving on Wednesdays, she will not be satisfied. But I have given her an answer supported by a reason: the prohibition itself. Where the cancelling permission does not apply, a prohibition does apply—a relatively well-defined prohibition, in fact. It is enough to infer a prohibition from a permission that the reasons for an action and the reasons against it share a practical context. To put this another way, a permission operating outside a prohibition makes no sense.

Because this is how permissions work, we can concede that a justification defense is a cancelling permission, that a cancelling permission is a second rule relative to the prohibition, and that the two rules are opposed to each other. The discontinuity of offenses and justification defenses does not follow from this. The cancelling permission—the justifi-

cation defense—is properly said to be continuous with the offense in the sense that it never operates outside the offense.

3. Prohibitions, Defenses, and the Specification of Ends

Finkelstein's description of an exception to a legal rule supports the continuity view as a substantive matter. One point of clarification at the outset of this argument is necessary. Finkelstein is right to say that an exception arises when we have a positive reason to recognize one.[159] She describes this conflicting reason as a principle. I think the conflicting reason is more often a conflicting end. Principles govern how we pursue our ends or which ends we pursue. An exception to a rule might arise when, for some reason, we ought to be governed by two conflicting principles in the pursuit of some end. But this seems like a rare occurrence. More often, it is our ends that conflict; we wish to attain or accomplish two incompatible things. An exception arises more often, and in more interesting ways, when we have conflicting ends. This, in any event, is how I will employ Finkelstein's insight.

We are accustomed to thinking that we reconcile conflicting ends by balancing, but conflicts of ends can be resolved other than by balancing. In fact, the whole idea of balancing ends is misleading. To say that we balance competing ends suggests that at some point we choose one of them as the dominant end and carve out some aspect of the losing end in order to effect the proper balance.

We would do better to think of reconciling our ends as a process of reciprocal specification of conflicting ends.[160] One of the principal features of Aristotelian ethics is the notion of deliberation on ends and an appreciation for the role such deliberations have, not only in constituting virtue but also in determining the individual acts that contribute to virtue.[161] The specification of ends is one mode of deliberation on ends. I might pursue health as an end, but if I hope to achieve it, I will have to specify that end in particular choices.[162] I will choose more immediate ends, such as getting adequate sleep or eating better food. These are means to health, and a means-ends relationship is one kind of specification. But notice that the means are also ends. In order to get better sleep, I will buy a comfortable bed. In order to eat well, I will join an organic food cooperative. To reflect this feature of deliberations on ends, let us abjure the terminology of means and ends, and refer instead to trailing and leading ends.

159. *Id.* at 515.

160. *See* HENRY RICHARDSON, PRACTICAL REASONING ABOUT FINAL ENDS 69–74 (1994); T.M. Scanlon & Jonathan Dancy, *Intention and Permissibility*, 74 PROC. OF THE ARISTOTELIAN SOC., SUPPLEMENTAL VOLUME 301, 308 n.8 (2000).

161. *See* RICHARDSON, *supra* note 160, at 54–57.

162. *See id.* at 70 ("Specifying an end will involve spelling out the ways or circumstances in which it is to be pursued, and the like.").

Notice that leading and trailing ends have no fixed referents. Health can be a leading end, as it was in my first example, but of course health can be a trailing end as well. I pursue health in order to succeed in my career. I pursue career success to achieve wealth and fame. I pursue wealth and fame to satisfy my parents' aspirations for me. And the sequence continues indefinitely. It might be that this regress ends with a final end—the good, or justice, for example—but we can leave that aside. For now, notice one last, obvious thing about ends, which is that they often conflict. My pursuit of success in my career might be hard to manage if I also want to be a devoted husband and father, and my pursuit of success might even conflict with my pursuit of health, which is, at other times, a means to my success.

An important feature of deliberations on ends, then, is the resolution of these conflicts, and the specification of ends is one way in which ends are reconciled. Often a trailing end is pursued in a particular mode in response to a particular mode of the leading end. A healthy life is the life of an athlete or the life of a non-smoker, depending on one's reasons for seeking health. A successful career is a high-paying career or an altruistic one depending on one's idea of success. We reconcile conflicts between ends when we specify them; that is, when we choose the particular mode of the trailing end that reflects the influence of the leading end. If I include family leave and at-work child care in my conception of a successful career, I can more easily maintain both career success (the trailing end) and success as a father (the leading end) in my scheme of ends. And because trailing ends have no fixed referents, the resolution of conflicts in ends is often a matter of reciprocal specification of ends. We take what was the trailing end, treat it as the leading end, and seek a mode of what was the leading end that will reconcile further conflicts. If I choose to be the father of one instead of the father of seven, then fatherhood (the trailing end) and career success (the leading end) are likely to conflict less often.

Our choices of elements in the drafting of offenses and justification defenses specify ends in this way. The criminal law is quite obviously concerned with the conflicting ends of autonomy and safety. We can take autonomy initially to be absolute; that is, we are free to kill another person just as we are free to do anything else we choose to do. But of course this complete autonomy is quickly stopped short by our conflicting interest in our own safety, which requires some measures to keep the peace. We specify the end of autonomy so that it excludes killing at will. The end of autonomy so specified, however, will specify the end of safety in turn. We will not prevent killing at will by peremptorily imprisoning anyone who seems to threaten harm. From this crudely defined starting point, we will pursue both autonomy and safety through their reciprocal specification along several different lines, down to the fine details of criminal law.

One line of the reciprocal specification of safety and autonomy in criminal law concerns the rule of law. Criminal prohibitions serve the end of safety, broadly speaking, but they will be closely defined ex ante by legislation because arbitrary punishment detracts from autonomy more than predictable punishment does. Even so, we will not aspire to perfect clarity or perfect notice in drafting criminal prohibitions, or attempt to eliminate all discretion in their enforcement. We will do this in order to close the loopholes inherent in a regime of rules, for the sake of safety.

Another line of safety-autonomy specification in criminal law concerns desert. For the sake of safety, we will prosecute and punish those who violate criminal prohibitions. But we will require proof of fault in this process because undeserved punishment is gratuitous and cruel, and because such punishment impinges on autonomy. In the proof of fault, we will sometimes use non-intentional fault criteria, because intentional fault criteria require the acquittal of idiosyncratic believers—who are no less dangerous for that. Safety is the leading end. But at the same time, we will use intentional-states fault criteria whenever we can because non-intentional-states fault criteria grant too much discretion to those who authorize punishment, detracting from autonomy.

A third line along which safety and autonomy are specified concerns the stringency of the criminal law. We will acquit an actor who is mistaken about the results of his conduct or the circumstances in which he acted because mistakes are an inevitable product of the exercise of autonomy. We will not allow mistake of law or willful ignorance as a defense, however, because to do so would encourage loopholing, impinging on safety.

The idea of reciprocal specification of ends explains why justification defenses are continuous with offenses even if, as Finkelstein's insight suggests, justification defenses serve ends that conflict with the ends served by the offense. These cancelling permissions are specifications of legal prohibitions. In other words, the definition of offenses and justification defenses occurs along continuous lines of specification that pass unbroken from offense to defense. Neither our rule-of-law concerns, nor our requiring proof of fault, nor our interest in the stringency of law lapses between offense and justification defense. The specification of safety and autonomy along these three discrete lines constitutes continuity between offense and justification defense.

I have already noted, in connection with criminal fault in provocation and mistake, two other conflicting ends that we pursue along lines of specification leading to the justification defenses. These ends are formality on one hand and fine-grainedness on the other. Because legal punishment involves the deliberate infliction of pain and even death, we want criminal law to be formal, in order to reduce arbitrariness and serve the other rule-of-law values. But because legal punishment involves the

deliberate infliction of pain and even death, we do not want our legal judgments to be over- or underinclusive relative to our moral judgments. We want fine-grainedness in these judgments, which obviously conflicts with formality. As I have argued, the interplay of intentional and non-intentional fault criteria reflects this conflict between the ends of formality and fine-grainedness. This conflict and its resolution by specification can be seen in the definition of offenses and justification defenses, respectively.

Our rule-of-law values strongly predominate in the definition of offenses, in that formality is most often the leading end, specifying fine-grainedness (such as we find in corresponding moral prohibitions) as a trailing end. For example, malice murder or depraved heart murder give us fine-grained results, but these definitions are clearly in disfavor. The Model Penal Code tied malice and "depraved heart" to an intentional-states fault criterion in the "recklessness with extreme indifference to the value of human life" formulation.[163] The drafters did this to increase the formality of these offense definitions and reduce the discretion of the jury to produce fine-grained results in murder cases.[164] The drafters chose a mode of the trailing end, fine-grainedness, that would serve the leading end, formality.

In justification defenses, fine-grainedness takes over. Formality, predominant in the definition of the offense, is now taken as the trailing end and is specified by the leading end of fine-grainedness. This specification is most obvious in the defense of lesser-evils. The lesser-evils defense is a formal norm when compared to its moral counterpart, but it is written in terms of standards, not rules, as an offense is usually written, because the defense of lesser evils aims at fine-grained results.[165] This arrangement runs so strongly against the formality of offenses and our rule-of-law values that some jurisdictions refuse to recognize lesser evils at all.[166] Where this defense is recognized, however, the end of formality is quickly given back its leading role for further specification of fine-grainedness. The choice of lesser evils as a matter of fact is left to the defendant, but only subject to the court's choice of lesser evils as a matter of law.[167] In this way, a process that bears on fundamental values in crim-

163. MODEL PENAL CODE § 210.2(1)(b) (1962).

164. *See id*. § 210.2(1)(b) cmt. 4 (noting that common law formulations of extreme indifference murder impose "great cost in clarity" and impair the "primary purpose of communicating to jurors in ordinary language the task expected of them").

165. *See id.* § 3.02 cmt. 3 (rejecting "any limitations on necessity cast in terms of particular evils to be avoided or particular evils to be justified").

166. In Canada, the lesser-evils formulation of necessity was rejected because to "hold that ostensibly illegal acts can be validated on the basis of their expediency would import an undue subjectivity into the criminal law." Perka v. The Queen, [1984] 2 S.C.R. 232, 248 (Can.).

167. *See* MODEL PENAL CODE § 3.02 cmt. 2 (explaining that the defense "requires that the harm or evil sought to be avoided be greater than that which would be caused by the commission of the offense, not that the defendant believe it to be so").

inal law begins in the definition of offenses and continues through the definition of justification defenses.

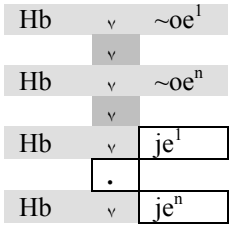
Notice that the "lesser evils as a matter of law" specification of fine-grained lesser evils for the sake of formality appeals to the offense in order to define the defense—another indication of continuity. For example, suppose the human resources director of a corporation intends to unilaterally shift all employees to a cheaper but grossly inadequate health insurance plan. The defendant might be correct in judging that to hold this executive hostage to force him to change his mind will, on balance, produce less suffering. Such a defendant will not be permitted to make this argument, however, because the law rules out such a choice—here, in the heavily rule-based prohibition on kidnapping. A case such as this one falls in the core of that offense.[168] The result is over- and underinclusion in lesser evils cases, relative to our moral judgments on some cases—the inevitable result of legal formality.

The other justification defenses can be seen as bearing the same relation to the defense of lesser evils. That is, these defenses, such as self-defense, are rule based in comparison to the standard-based lesser-evils argument because their definitions represent a specification of lesser evils' fine-grainedness by the (once again leading) end of formality. On the other hand, within self-defense and similar formal versions of lesser evils, the end of fine-grainedness again takes the leading role in the particular terms of these defenses—in the rules on reasonable mistake, for example. But for now, let us keep the focus on these defenses as specifying the fine-grained defense of lesser evils under the leading end of formality.

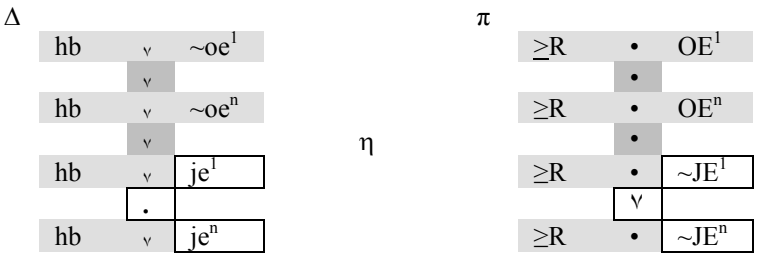
The specification of ends in the justification defenses accounts for the troublesome flip in connectors on the vertical axis. Consider the common, indispensible element of each of these more closely defined defenses: the requirement of necessity. Standing alone as a justifying consideration, necessity states the lesser-evils principle. If, in pursuing a claim of self-defense, the defendant were able to prevail only on a showing of necessity, we would have moved backwards in the process of specification, collapsing self-defense into lesser evils. And it would be a simplistic form of lesser evils at that, lacking the refining details of the formal version, such as the court's control over lesser evils as a matter of law. If we are to avoid this backsliding into the broad defense of lesser evils, we will have to require the defendant to make a prima facie case and raise a reasonable doubt about necessity *in conjunction with* the other elements of self-defense. The paired elements of self-defense and other justification defenses will be connected, vertically, by a conjunc-

168. *See id.* ("What is involved may be described as an interpretation of the law of the offense, in light of the submission that the special situation calls for an exception to the criminal prohibition that the legislature could not reasonably have intended to exclude, given the competing values to be weighed.").

tion—in contrast to the proof required of the defendant on the elements of an offense. The formality of this conjunction of elements in the specific justification defenses will specify the trailing end of fine-grainedness served by the lesser-evils argument. In graphic terms:



Adding the implicit proof required of the prosecution, we have:



This is, of course, the continuity view of criminal wrongdoing.

The function of this conjunction is the same in the lesser-evils defense itself. The conjunction of elements is needed to keep the lesser-evils defense from devolving into bare necessity and to maintain the defense of lesser evils in its fully elaborated form. Or, more to the point, conjunction of elements is needed to maintain the process and results of the specification of ends in justification defenses.

Now, in one sense this is obvious. Taking away the artificial diachronic language I have used in describing the specification of ends, we might say that, if not for the conjunction on the vertical axis on the defense side, self-defense and other formal justification defenses would be superfluous, or that the defense of lesser evils would be too broad. We do not need the idea of specification of ends to see this. We do, however, need the idea of specification of ends to explain it adequately, as something other than an isolated feature of the justification defenses. Once we explain the justification defenses with a specification account, we can see the conjunction on the vertical axis on the defense side as a step in a process that reaches back beyond the definition of the defense and into the definition of the offense. It is a product of and evidence for the continuity of criminal offenses and the justification defenses.

C. Overinclusion, Underinclusion, and Second-Best Solutions

The *Dadson* view is at least as objectionable as the continuity view and its implicit genuine justification thesis. The treatment of the mista-

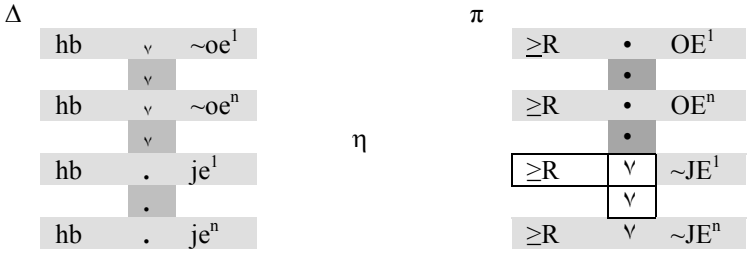
kenly justified actor has implications for the mistakenly unjustified actor. The detective will be convicted on the *Dadson* view, but so will the runner—who seems to deserve an acquittal. The runner will be acquitted on the continuity view, but so will the detective—who seems to deserve a conviction. But this is not to say that we have nothing more than competing intuitions to choose between. These counterintuitive implications can be articulated in more detail, and doing so can give us reasons to prefer one view over the other.

The principal difficulty with the *Dadson* view is that, because the prosecution can prevail by proving only fault regarding justifying circumstances, it results in overinclusion. That is, some cases that we would not condemn morally, such as the runner's, will be treated as crimes. The rogue detective has no redeeming beliefs, so to speak, and the *Dadson* view condemns him in accordance with our intuitions. But *Dadson* also condemns the otherwise entirely justified actor who holds a single, false belief about a non-justifying circumstance.

Suppose that the hitman is at the door of the victim's apartment, about to shoot and kill him. In this case, however, the detective is fully aware of the hitman's intentions and has come to attempt to prevent the murder. The intended victim is a former professional associate of the hitman who is now in a witness protection program. The detective has received a report that the witness's cover has been blown and that the hitman is on his way. The detective approaches the apartment door from such an angle that he can see the victim frozen in terror and the hitman raising his gun. The detective shoots and kills the hitman. In his debriefing by Internal Affairs after the incident, the detective says that it did cross his mind a moment before he fired that there might have been time to shout, "Stop! Police!" and that there was a genuine, though small, possibility that the hitman would comply. The Internal Affairs detectives follow up on this suggestion by asking the intended victim for his view on that possibility. He tells the detectives that, no, he had seen a look in the hitman's eyes that meant he was moving in for the kill and was beyond all recall. The victim had seen this look several times before when the original plan was merely to threaten someone, and the hitman had gone too far in spite of everything his cohort did to stop him. There was absolutely no question in the intended victim's mind that his death was imminent, and that nothing short of the hitman's own death would have prevented it.

Under *Dadson*, this conscientious detective will be convicted of the murder of the hitman. He consciously disregarded a possibility that the hitman might surrender—something that, as it turns out, was never a possibility. He will be able to show $je^1 \ldots je^n$ and also hb as to all but the immediate necessity of using force. He is guilty of murder, however, because he is reckless in this one respect. The missing non-justification elements are immaterial, given that the horizontal connectors are dis-

junctive. Similarly, the fact that the conscientious detective is at fault in only one respect is immaterial, given that the vertical connectors are disjunctive. Highlighting the prosecution's route to a conviction under the *Dadson* view, we can see that the commission of a complete murder rests on fault alone.



This is uncomfortably close to punishment for thoughts. If a belief in a single fact—one that did not, in fact, obtain—is all that stands between the conscientious detective and an acquittal, then he is punished for a belief, and not for wrongdoing. Granted, he is a murderer. All the offense elements and associated fault elements are present. This is not punishment for thoughts in the sense that no criminal act has occurred. But it is jarring to think of the conscientious detective as a murderer, and this cognitive dissonance matters. It tells us that the proof of the offense is moot once a justification defense has been raised, and that the question of criminal wrongdoing now turns entirely on the success or failure of that defense. If the defense fails solely because the detective consciously disregards a risk of something that never existed, he is truly punished for a thought alone.

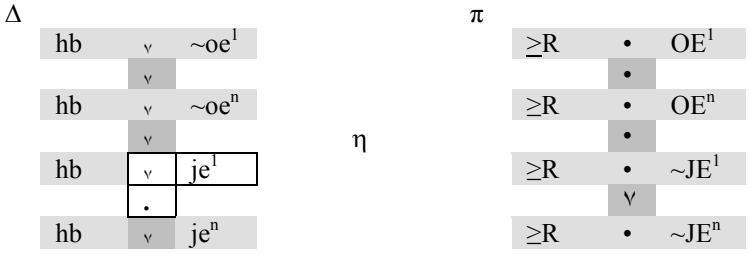
Defenders of the *Dadson* view are, of course, prepared with answers to this objection. The *Dadson* view has two second-best solutions in this case—two adjunct theories that avoid counterintuitive results, albeit with something less than a seamless extension of the main theory. The first argument is that the conscientious detective, like the runner, will be excused in a *Dadson* jurisdiction. This move suffers from the fact that it rests on only a putative excuse, for reasons that I explained in Part II. This kind of excuse is not a defense at all; it is a showing of nonfault in the commission of the offense. The upshot of this kind of "excuse" is the very problem that the proponents of the "excuse" solution are attempting to solve: a conviction that follows from the way in which they have connected the fault elements of the justification to the fact elements of the justification. Apparently, the courts and scholars who adopt the *Dadson* view have a different kind of excuse in mind. They have not said, however, what it is or how it works.

The second second-best solution is to convict the conscientious detective of an attempt. Had the situation been as he supposed—had the hitman been willing to surrender and not to kill—the detective would have committed an unjustified killing. It is no thanks to him that the

hitman was undeterrable and that the killing was, in fact, justifiable. The detective was apparently willing to commit a crime. This is the gist of attempt, and attempt is the appropriate resolution of this case. This move has the merit of addressing the allegation of punishment for thoughts. An attempter's liability can always be said to turn on his fault alone. The harm he sought to do did not come about, leaving only the willingness to make it so as a basis for criminal liability. We have no qualms about this—we do not think of it as punishment for thoughts alone—and nor should we see the conscientious detective's case that way.

The difficulty is that the *Dadson* view precludes attempt liability. A mistakenly justified actor cannot be guilty of an attempt to commit an offense if he is also guilty of the complete offense, under the principle of merger.[169] The absence of the non-justification elements, which would otherwise make the offense an attempt, does not forestall the crime's being a complete murder on the basis of the defendant's beliefs alone. To excuse the complete murder in order to get around the merger doctrine is, of course, not an option. Again, this is only a putative excuse because of the *Dadson* view itself. As a result, the *Dadson* view has no second-best solution to the conscientious detective's case. He is a murderer because he had a belief—that the hitman might possibly surrender—that was never true.

It almost goes without saying that the continuity view does not have this counterintuitive implication. Because the horizontal axes are conjunctive on the prosecution's side and disjunctive on the defense side, the actual immediate necessity in the case will suffice to acquit the conscientious detective. Here is the continuity view, highlighting the conscientious detective's route to this result.



All justification elements are present in the case of the conscientious detective, as are beliefs in each of them—except for one. In that pair, the justification element is decisive. The disjunctive horizontal connectors and the conjunctive vertical connector on the defense side allow one justification element—the fact of immediate necessity—to moot the effect of his missing belief in it.

169. *See* DRESSLER, *supra* note 81, at 408 ("If a person commits the target offense, she may *not* be convicted of both it and the criminal attempt.").

This does not suggest that the continuity view is never in need of a second-best solution of its own. Both kinds of closure view—negative elements and continuity—have their own counterintuitive implications. They acquit the rogue detective. The second-best solution for the continuity view is to convict the rogue detective of an attempt. In contrast to the *Dadson* view, however, the continuity view produces a viable attempt. That is to say, it *has* a second-best solution, even though, as we will see in Part V, it does not need one.

To fully appreciate the continuity view on this point, it is helpful to consider the negative elements version of the attempt gambit first. The rogue detective thought he had no justification for killing the hitman, but the situation was not as he thought it to be. He did have a justification for killing the hitman, in every respect. He tried, but failed, to commit murder, and should be guilty of an attempt. This highlighted table shows the prosecution's route to attempt liability, along with the connector that precludes the proof of a complete murder by the rogue detective.

Δ					π		
hb	∨	~oe^1			≥R	•	OE^1
	∨					•	
hb	∨	~oe^n			≥R	•	OE^n
	∨			η		•	
hb	∨	je^1			≥R	•	~JE^1
	∨					•	
hb	∨	je^n			≥R	•	~JE^n

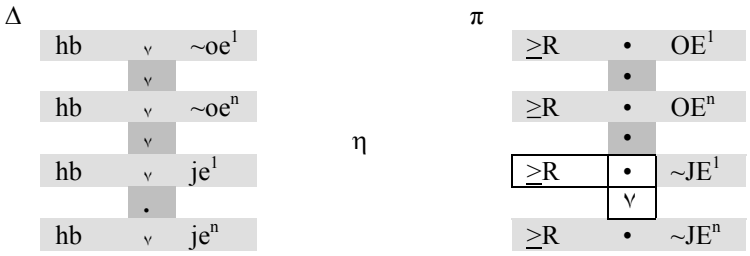
The detective's willingness[170] to act without justification—his fault with respect to the justifying circumstances—is sufficient for the commission of an attempt. Given that the horizontal connectors are "•" on the prosecution side, we do not have a complete murder, avoiding the merger that is fatal to the *Dadson* view's attempt gambit. On the negative elements view of wrongdoing and justification, the attempt gambit appears to work.

The negative elements view is clearly wrong on the vertical axis, however, and this is fatal to its version of attempt as a second-best solution to the problem of the rogue detective. Some cases that are indistinguishable from the detective's case, morally speaking, but which we would also wish the law to capture as attempts, are not attempts under the negative elements view. They result in acquittal, the counterintuitive result that the second-best solution is meant to avoid.

170. The Model Penal Code prescribes different levels of requisite fault to three different kinds of element—conduct, result, and attendant circumstances. The elements of justification defenses are usually attendant circumstances, and for the most part these elements retain the same requisite fault that they have in the definition of the complete offense. I need to elide these complications for the sake of a general description, and so will use the words "willingness" or "choose" to refer to the requisite fault.

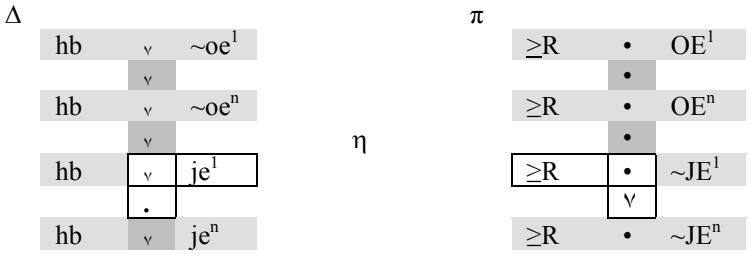
Under the negative elements view, the vertical connectors are conjunctive all the way down on the prosecution side. This implies that a willingness to act without justification in every respect is necessary to prove an attempt, just as a willingness to commit all elements of the offense is necessary to prove an attempt. The rogue detective from the original hypothetical meets these conditions and has committed an attempt. But now take the case of one whom I will call the eager detective. As in the original version, all justification elements are present. And as before, the detective is unaware that the threat is imminent or that deadly force is necessary to avert it. Now, however, the detective knows that the apartment is occupied and that the hitman intends to kill the occupant. This means that justifying circumstances are present in every respect, but he does not choose to act with justification in every respect. That is, he is aware of a threat of death to another, but for all he knows the threat might be averted without the use of any force at all. On the negative elements view he has not, therefore, committed an attempt, even though he knows he could have stopped the hitman some other way and is, it seems, attempting to kill without justification. The prosecution cannot prove that he was willing to act without justification in every respect, as the negative elements view requires by virtue of its vertical conjunction of the paired belief and justification elements. But by the same token, the eager detective has not committed a complete offense, as we have just seen. He will be acquitted of both the complete offense and the attempt.

The continuity view presents a viable attempt because it is not underinclusive in this way. That is, there are no cases that are indistinguishable from the rogue detective's case, morally speaking, that the law will fail to capture as attempts. On the continuity view, the vertical axis on the proof of justification defenses is disjunctive on the prosecution's side—the opposite of the vertical axis in offense definitions. This means that a purpose to act without justification in every respect is not necessary to prove an attempt, even though, as always, a purpose to commit all elements of the offense is necessary to prove an attempt. Suppose again that the detective knows that the apartment is occupied and that the hitman intends to kill the occupant, but he does not know that the threat is imminent or that deadly force is necessary to avert it. The eager detective will not be guilty of a complete crime because all justification elements are present—implying, of course, that all non-justification elements are missing. On the continuity view, the showing of fault with respect to any of these non-justification elements will suffice to establish the attempt. This is so because, unlike the negative elements view, the vertical connector between justification elements on the continuity view is a disjunction, "∨." Here is a version of the continuity view showing the route to this result.



The horizontal "•" in the prosecution's case forestalls conviction for the complete crime. That is, it allows the defendant to succeed in showing justification, avoiding the merger that would occur under the *Dadson* view. The vertical "∨" in the prosecution's case facilitates conviction for the attempt, because it renders the honest belief in the existence of a threat moot, so far as attempt liability is concerned. That is, on the negative elements view this honest belief would preclude attempt liability because the prosecution could not prove that the eager detective was at fault with respect to every non-justification element, as the negative elements view requires. Under the continuity view, the vertical "∨" allows us to ignore the honest belief in the existence of a threat and obtain an attempt conviction based on any aspect of the case regarding which the detective was willing to kill without justification. The fact that the eager detective did not believe that death was imminent, for example, would suffice for attempt liability. The continuity view produces a viable second-best solution to mitigate its acquitting the rogue detective. It produces an attempt not only in that case, but also in the case of the conscientious detective, which the *Dadson* view underincludes, and in the case of the eager detective, which the negative elements view underincludes.

One problem remains, however. The conscientious detective is in the same position as the eager detective on the continuity view. I gerrymandered the facts to make one actor appear more appealing than the other, but both are willing to kill while aware of a risk, at least, that it is not necessary to do so. In graphic terms, both of them have this route to an acquittal for the complete crime and a conviction for an attempt.



The disjunctive horizontal connectors and the conjunctive vertical connector on the defense side allow one justification element to moot the effect of the missing belief in it, resulting in an acquittal for the complete

offense. On the prosecution side, the conjunction on the horizontal axis precludes liability for the complete offense, thus avoiding merger, while the disjunctive vertical axis makes his fault in only this one respect sufficient for an attempt.

If one balked at the acquittal of the eager detective on the ground that he is genuinely justified, this is no reason to reject the continuity view. The eager detective can be convicted of an attempt, as can the conscientious detective. On the other hand, if it seems that the conscientious detective should be fully acquitted while the eager detective should be held to have committed an attempt, then the continuity view seems to have produced a new dilemma in its bid to offer a second-best solution to the original one. However, this apparent dilemma is a product only of the facts of individual cases and not of the underlying structure of elements. That is, we have only a fact-dependent choice between full acquittal and attempt liability under the only view that produces a viable attempt at all. This is not an easy choice, but it is not a dilemma over how to structure the law of justification defenses.

V. GENUINE JUSTIFICATION AND THE LAST SPECIFICATION

Even if we assume that the counterintuitive implications of the continuity view can be mitigated by our finding the mistakenly justified actor guilty of an attempt, this successful second-best solution is morally indefensible.

Attempt liability has a troubling feature at its core. If the punishment imposed for an attempt is less than it is for the complete offense—as it usually is[171]—the defendant receives the benefit of a fortuity.[172] The crime is an attempt because he has not inflicted harm, but this is no thanks to him. In the case of the mistakenly justified actor, however, we have a similar problem that is much, much worse. If we impose only attempt liability on a killer such as the rogue detective, he receives the benefit of a fortuity. But the fortuity is not that no harm has been inflicted—the victim is dead—it is simply that the applicable rules must be written so as to benefit the mistakenly justified actor if we are to find a mistakenly unjustified actor, such as the runner, innocent. The fortuity, in other words, is a fortuity of legal formality. There is no saving grace in this, as there is in the absence of harm in an ordinary attempt. The value of formality in law is beyond measurement, and it is profoundly dangerous to dismiss any formal feature of law as a "mere technicality" that ought to give way to moral judgment or to the underlying balance of reasons. Even so, we find it difficult—understandably, I think—to see formality as a saving grace when we acquit, or mitigate the liability of, a killer such as the detective, and no less so in less serious cases.

171. *See* DRESSLER, *supra* note 81, at 415.
172. *See id*. at 417.

My thesis has been that the mistakenly justified actor is genuinely justified, and I think that I have demonstrated that he is. However, I have not said that this is a morally acceptable, or even a defensible outcome, and I do not believe that it is. The question now is whether we need to accept it. We do not. We face a dilemma, but it is important to see what kind of dilemma it is. It is a legal dilemma, not a moral dilemma. Our moral judgments on the two kinds of mistake in justification cases are clear and consistent. The mistakenly unjustified actor should not be punished, and the mistakenly justified actor should be punished. It is our legal judgments that are at odds: we must convict both or acquit both. The problem, as I have said, is legal formality, and this calls for a formal solution. We need to change the law.

Changing the law to avoid a particular outcome is problematic. Result-oriented law making is unprincipled, almost by definition—but only almost. A change in the law can be result oriented as a matter of motivation if it is guided by principle in its execution. If we are guided by the same principles that brought us to the dilemma, carrying them one step further to resolve it is hardly objectionable. To solve our dilemma concerning mistakes in justification, then, we need one last specification of the ends involved in the justification defenses. As it happens, we can identify at least two distinct lines of specification that support the solution I propose here.

To see these specifications and their significance, let us begin with the rule that would be added to existing law to serve them. The rule is best described as a limited, conditional imposition of the *Dadson* view. Let us call it the qualified continuity view. *If all justification elements are present as a matter of fact, then the defendant must present evidence of an honest (reasonable) belief in each and every justification element, sufficient to raise a reasonable doubt about fault with respect to each and every non-justification element.* Put more succinctly in terms of horizontal and vertical connectors, the rule is that, for the mistakenly justified actor only, if and only if all justifying elements are present, the connectors on the horizontal axes on the defense side are conjunctive, not disjunctive. In all other cases, the ordinary continuity view would prevail.

The reason for the limited application of the *Dadson* view to the mistakenly justified actor, but not to the mistakenly unjustified actor such as the runner, is easy to see. The continuity view suffices to acquit the runner, matching our moral judgment of her case. The reason for the conditional application of *Dadson* is also not difficult to see. The vertical connector is disjunctive on the prosecution side, on both the *Dadson* and the continuity views—as the overwhelming weight of authority says it must be. This means that if the defendant cannot show imminence in self-defense, for example, then no matter how we arrange the connectors on the horizontal axes, this missing justification element will doom his defense. The problem arises only when the vertical axis is not sufficient

to convict an actor such as the rogue detective—that is, when all justification elements are present.

This rule has an unmistakable odor of the ad hoc about it. It tailors the continuity view to capture the mistakenly justified actor. Or, to put it another way, it is *Dadson* as *deus ex machina*. But we can see that this objection is no objection at all if we frame it a third way: the rule is formalistic, in the sense that it elevates form over substance. The answer to the objection in this light is that, in this case, form is substance. Our problem with the rogue detective arises because formality overbears our moral judgments about these cases. Like all attempters, he benefits from a fortuity—which in his case is not non-harm, but the formal rules governing the fault and fact elements of the justification defenses. In spite of the value of legal formality generally, this seems like a flimsy reason not to convict the mistakenly justified actor of the complete crime. To say that the law fails to match our moral judgment about the case is to say precisely that the law is insufficiently fine-grained. And if this failure is occasioned by what seems like excessive formality, a final round of specification of formality by fine-grainedness seems justified.

In the definition of offenses, formality is most often the leading end, specifying fine-grainedness as a trailing end. The justification defenses are a specification of that formality, now taken as the trailing end, by fine-grainedness. In the end, we saw that any backsliding into too much fine-grainedness is stopped by the formal requirement that the elements of the justification defenses such as self-defense must be proved conjunctively on the vertical axis (from the defense point of view), so that necessity standing alone does not obviate the other elements. A further specification of formality by fine-grainedness, not reflected in existing law, would rule the mistakenly unjustified actor within, and the mistakenly justified actor out of, the category of genuine justification. This round of specification produces the qualified continuity view. This limited application of *Dadson* is a rule that aims at fine-grainedness, and not merely at the mistakenly justified actor.

The second line of specification that supports the qualified continuity view concerns the ends of autonomy and safety. Autonomy is one end of any criminal law defense, simply because it limits the reach of criminal law into individuals' lives. Its relevance to justification, however, goes well beyond this, to autonomy-based considerations that are not to be found in non-responsibility defenses. The non-responsible actor is in the degrading position of asking not to be punished because he is less than a full, equal participant in society. In contrast, a justification defense affirms human dignity because it affirms moral agency. If the accused chose the lesser evil correctly, he will be acquitted, but whether his act was in fact the lesser evil is a question committed to the jury with very

little to constrain its decision.[173] The actor is at risk of being second-guessed in hindsight, but that risk is entailed in a grant of responsibility—a grant of power carrying consequences, both for those affected by the decision and for the decisionmaker. This grant of responsibility in the lesser-evils defense reflects the end of autonomy.

In the next round, the specific justification defenses such as self-defense reassert safety as the leading end by limiting the sweep of the lesser-evils principle, thus qualifying the autonomy implicit in the lesser-evils defense. And in the round after that, the fault elements in these defenses, such as unreasonableness, again put autonomy in the leading role. It is at this point that we encounter the dilemma of our either acquitting or convicting both kinds of mistaken actors, without distinction—contrary to our moral judgments about each of them. A further specification, this time with safety as the leading end, is called for. The rule of the qualified continuity view serves this purpose. A limited, conditional application of *Dadson* specifies our fault-governed definitions of the specific justification defenses in a way that serves the end of safety. We will convict the mistakenly justified actor if all non-justification elements are present, but only if he is at fault with respect to each of them. We will not apply this rule to the mistakenly unjustified actor, however, effectively leaving her case at the previous round of specification that serves autonomy, by acquitting her based on her judgment about the situation, as she saw it.

VI. CONCLUSION

In this Article, I have pursued a legal theory of legal justification. I have examined the elements of offense definitions and justification defenses—or, more accurately, the conjunctions and disjunctions between these elements—to resolve a longstanding issue in the theory of legal punishment. I have shown a principled way to rule the mistakenly unjustified actor within, and the mistakenly justified actor out of, the class of justified actors, thus matching our legal judgments to our moral judgments concerning these cases.

In the end, however, I think the route to this conclusion is more important than the conclusion itself. The tables of elements and their connectors are illuminating—surprisingly so for such a simple device. This elements analysis enables us to demonstrate some elementary points, such as the existence of a genuine difference between offenses and justification defenses. And it permits us to demonstrate more esoteric points, such as the unavailability—on any view of wrongdoing and justification—of attempt liability as a second-best solution to the dilemma presented by mistakes in justification.

173. *See* MODEL PENAL CODE § 3.02 cmt. 4.

I have used Aristotelian punishment theory as a complement to elements analysis. The use of virtue ethics in my analysis did not violate my pledge to pursue legal theory and not moral theory, because I did not simply equate solutions to moral issues with solutions to legal issues. Instead, I used a method of Aristotelian analysis—its conception of deliberations on ends as specification of ends—as a model for legal analysis. This enabled me to model the justification defenses as the product of the specification of competing ends—making the case that the justification defenses, even taken as permissions that cancel the prohibition for reasons that conflict with the prohibition, are continuous with criminal prohibitions.