THE YALE LAW JOURNAL

GREGORY ANTILL

Fitting the Model Penal Code into a Reasons-Responsiveness Picture of Culpability

ABSTRACT. This Note compares the Purpose, Knowledge, Recklessness, and Negligence (PKRN) mens rea regime laid out in the Model Penal Code (MPC) and dominant in American criminal law with the "reasons-responsiveness" conception of culpability widespread among contemporary philosophers and criminal-law theorists. Whereas a PKRN picture of culpability sorts an agent's culpability for an action according to whether the action was performed purposefully, knowingly, recklessly, or negligently, the reasons-responsiveness picture locates an agent's culpability in the responsiveness of the agent's reasoning capacities, which their actions evince. While many criminal-law theorists are cognizant of these different conceptions of culpability, most have assumed that the two pictures of culpability generally converge when it comes to the relative culpability judgments of actions performed purposefully, knowingly, recklessly, or negligently, so that even for those who reject an underlying PKRN conception of culpability in favor of an alternative reasons-responsiveness conception, a PKRN mens rea regime can still provide a roughly adequate system of culpability proxies.

In contrast, this Note argues that criminal-law theory has deeply underestimated the degree to which these conceptions are in tension with one another. If the reasons-responsiveness picture of culpability is correct, we should expect frequent cases of interhierarchical disagreement between the reasons-responsiveness picture and the MPC grading regime. That is, we should expect frequent cases where, for example, an agent who acts purposefully is less culpable than an agent who performs that same action recklessly or negligently. This result has both important normative and empirical consequences for the practice and study of substantive criminal law. In particular, this Note argues that if the reasons-responsiveness account of culpability is correct, then the MPC's grading system will often fail to track offenders' relative culpability and result in predictably disproportionate punishments not merely within but also across grades of crimes.

AUTHOR. Yale Law School, J.D. expected 2023; Ph.D., University of California, Los Angeles, Philosophy. I'm grateful to the audience at the Yale Law & Philosophy Work-in-Progress Workshop, where an earlier version of this material was presented. Special thanks to Gideon Yaffe, Dan Kahan, Gabe Mendlow, Pamela Hieronymi, Kenneth Simons, Steve White, Lee-Ann Chae, Brian Hutler, Yuan Yuan, Pinchas Huberman, D Black, Joel Sati, and David Emer. I would also like to thank the editors at the *Yale Law Journal*, especially Thaddeus Talbot.

NOTE CONTENTS

INTRODUCTION

According to the familiar Purpose, Knowledge, Recklessness, and Negligence (PKRN) mens rea regime introduced in the American Law Institute's Model Penal Code (MPC) in 1962 and now dominant in American criminal law,[1] the criminal liability of an agent who commits some offense varies depending on whether the agent acted purposefully, knowledgeably, recklessly, or negligently with respect to the material elements of that offense.[2] MPC section 2.02(1) establishes that for a person to be guilty of an offense, they must have at least one of these four culpable mental states.[3]

MPC section 2.02(5) establishes a weak ordering among the four mental states. For any given material element, each of the four PKRN mental states involves as much or more liability than the subsequent state.[4] Where the severity of crimes is graded based on mens rea, this ordering hierarchy is typically used to establish the respective grades.[5] If, for example, someone *causes* the death of another person (the material element of criminal homicide under MPC section 210.1), their criminal homicide will be classified as a murder if their action was purposeful. But they will typically be guilty merely of manslaughter, a lesser offense, if they were only reckless with respect to the victim's death, and of the even less severe crime of negligent homicide if they were only negligent with respect to the victim's death.[6] That is, if they did not intend the death they

1. *See generally* Paul H. Robinson & Markus D. Dubber, *The American Model Penal Code: A Brief Overview*, 10 NEW CRIM. L. REV. 319, 319-20, 326-29 (2007) (showing that over two-thirds of states have adopted the Model Penal Code (MPC) in whole or in part, and that "even within the minority of states without a modern code, the Model Penal Code has great influence").

2. *See* MODEL PENAL CODE § 2.02(1)-(2) (AM. L. INST. 1962). One major innovation of the MPC was to adopt an "element analysis" of criminal offenses, decomposing criminal offenses into various material elements including actions, the causal results of those actions, and attendant circumstances, *id.* § 1.13(9)-(10), along with corresponding mens rea elements for every material element, *id.* § 2.02(1).

3. *See id.* § 2.02(1). There are some exceptions for certain strict-liability crimes. *Id.* § 2.02(5).

4. *Id.* § 2.02(5).

5. *See id.* § 2.02(10). While the MPC regime always respects the weak ordering of purpose knowledge recklessness negligence, certain grading schemes treat neighboring mental states in the hierarchy as equivalent. For example, the MPC homicide grading regime treats purpose and knowledge as equivalent. *Id.* § 210.2(1)(a). Indeed, for purposes of determining guilt, the default for statutes without explicit mens rea grading is to treat purpose, knowledge, and recklessness as all equally sufficient to establish culpability. *See id.* § 2.02(3). A similar ordinal ranking exists in non-MPC regimes that follow the alternate "Penn System" of homicide grading. *See* Edwin R. Keedy, *History of the Pennsylvania Statute Creating Degrees of Murder*, 97 U. PA. L. REV. 759 (1949).

6. *See* MODEL PENAL CODE § 210.1-.4 (AM. L. INST. 1962).

caused (and so did not act purposefully under MPC section 2.02(a)), but merely believed there was some substantial probability that a death might result from their actions (and so acted recklessly under MPC section 2.02(c)), they will be considered less liable for that act, and so guilty of a lesser grade of offense and subject to lesser criminal penalties.

Influenced by historical mens rea distinctions from the old common-law homicide doctrine, the MPC does occasionally engage in more fine-grained parsing of mens rea in its analysis of criminal homicide than it does for other crimes, as with the addition of "reckless[ness] . . . manifesting extreme indifference to the value of human life"[7] (equivalent to ordinary purpose or knowledge for the purposes of grading under MPC section 210.2(b)), and purpose "under the influence of extreme mental or emotional disturbance"[8] (equivalent to ordinary recklessness for the purposes of grading under MPC section 210.3(b)). It also provides affirmative defenses that can function as full or partial shields to liability based on more finely grained features of the defendant's subjective psychology.[9] Still, despite the potential for such mitigating or aggravating factors to affect criminal liability on the margins, the PKRN hierarchy provides the backbone of the MPC's model of criminal liability and ensures at least a weak ordinal sorting of criminal liability. Though a purposeful homicide, for example, may sometimes be treated similarly to particularly severe cases of reckless homicide, the offender who commits a purposeful homicide will never be held less liable than an offender who commits negligent criminal homicide, nor will a case of aggravated reckless homicide be treated as involving more liability than ordinary cases of purposeful homicide.[10]

Underlying the MPC grading regime appears to be a crucial normative commitment to (1) the view that an agent's responsibility for some act, and hence their subjective culpability, is a function of the proximate mental states behind the act, and (2) a substantive view about which proximate subjective mental states are normatively *worse* (that is, make the agent more culpable) than others. The same agent performing the same act is more culpable if they intended the

7. *Id.* § 210.2(1)(b).

8. *Id.* § 210.3(1)(b).

9. See, for example, the affirmative defenses of necessity, *id.* § 3.02(1)(a), duress, *id.* § 2.09, or self-defense, *id* § 3.04. As discussed in Section I.D, *infra*, these features of the Code provide evidence that the MPC was itself influenced (at least implicitly) by the force of the reasons-responsiveness account of subjective culpability.

10. For additional discussion of the limitations of affirmative defenses and additional mens rea categories for increasing or decreasing criminal liability in the MPC, see *infra* Section I.D.

effect (purpose)[11] than if the effect was foreseen but unintended (knowledge),[12] even less culpable if merely a risk of the effect was foreseen (recklessness),[13] and less culpable still if they unreasonably failed to foresee any risk of the effect at all (negligence).[14] Call the conjunction of (1) and (2) the PKRN picture of subjective culpability. While this picture of subjective culpability provides the most straightforward explanation for the MPC's mens rea hierarchy, it is rarely articulated explicitly, and even more rarely defended.[15] Recent empirical work also calls into question the widespread assumption that the PKRN picture of subjective culpability maps onto the common-sense intuitions of the average lay juror.[16]

The MPC grading regime's apparent reliance on the PKRN picture of subjective culpability is particularly problematic because that picture is widely rejected by most contemporary moral philosophers and criminal-law theorists who hold instead what might broadly be labeled a "reasons-responsiveness" conception of culpability.[17] According to this reasons-responsiveness conception of

11. MODEL PENAL CODE § 2.02(2)(a) (AM. L. INST. 1962).

12. *Id.* § 2.02(2)(b).

13. *Id.* § 2.02(2)(c).

14. *Id.* § 2.02(2)(d).

15. For a qualified defense, see R. A. DUFF, INTENTION, AGENCY AND CRIMINAL LIABILITY: PHILOSOPHY OF ACTION AND THE CRIMINAL LAW (1991); and Kimberly Kessler Ferzan, *Don't Abandon the Model Penal Code Yet! Thinking Through Simons's Rethinking*, 6 BUFF. CRIM. L. REV. 185 (2002), which identifies a coherent "choice conception" of culpability that Kimberly Kessler Ferzan takes to implicitly underlie the MPC picture. *But see* John Gardner & Heike Jung, *Making Sense of Mens Rea: Anthony Duff's Account*, 11 OXFORD J. LEGAL STUD. 559 (1991) (questioning whether R. A. Duff's use of mens rea terms like "reckless" or "intentional" map onto their usage in the MPC).

16. *See* Francis X. Shen, Morris B. Hoffman, Owen D. Jones & Joshua D. Greene, *Sorting Guilty Minds*, 86 N.Y.U. L. REV. 1306, 1339-43 (2011) (observing that test subjects could not reliably distinguish between knowing and reckless conduct); *see also* Matthew R. Ginther, Francis X. Shen, Richard J. Bonnie, Morris B. Hoffman, Owen D. Jones, Rene Marois & Kenneth W. Simons, *The Language of Mens Rea*, 67 VAND. L. REV. 1327, 1351-53 (2014) (observing that test subjects had difficulty sorting mental states by culpability level in the order prescribed by the MPC).

17. These various views, which I will group under the rubric of "reasons-responsiveness accounts," encompass both what contemporary theorists call "quality of will accounts" and "reasons-responsiveness accounts." Contemporary philosophical proponents of reasons-responsiveness, broadly construed, include: NOMY ARPALY & TIMOTHY SCHROEDER, IN PRAISE OF DESIRE (2013); Pamela Hieronymi, *Responsibility for Believing*, 161 SYNTHESE 357 (2008); T. M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, AND BLAME (2008); Angela Smith, *Responsibility for Attitudes: Activity and Passivity in Mental Life,* 115 ETHICS 236 (2005); JOHN M. FISCHER & MARK RAVIZZA, RESPONSIBILITY AND CONTROL: A THEORY OF RESPONSIBILITY (1998); and SUSAN WOLF, FREEDOM WITHIN REASON (1990). Among criminal-law

culpability, an actor's bad state of mind consists not in their intentions, purposes, knowledge, or negligence, but rather in the responsiveness of their reasoning capacities, which their actions (given their purposes, knowledge, recklessness, or negligence) evince.[18]

This Note compares the PKRN mens rea regime with the reasons-responsiveness conception of subjective culpability. While many criminal-law theorists are cognizant of these different conceptions of subjective culpability, criminal-law theory has deeply underestimated the degree to which these conceptions are in tension with one another, and so underappreciated the downstream normative consequences of these two different underlying pictures of subjective culpability for substantive criminal law.

This Note proceeds in three parts. Part I explains in more detail the reasons-responsiveness conception of culpability and draws out the tensions caused by contemporary criminal-law theory's joint commitments to both a PKRN system of criminal liability and a reasons-responsiveness conception of criminal culpability.

Part II, the heart of this Note, argues that the degree of tension between these joint commitments has been underestimated. In particular, while criminal-law theorists have noted ways in which a reasons-responsiveness picture of subjective culpability may lead to more fine-grained distinctions of criminal responsibility than the PKRN picture, and so lead to what I call *intra*hierarchical differences in criminal responsibility for particular offenses, almost all theorists have assumed that the two pictures of subjective culpability generally converge when it comes to the general relative culpability judgments of actions performed purposefully, knowingly, recklessly, or negligently. These theorists therefore conclude that, even if we reject the PKRN picture of subjective culpability, we can still maintain the PKRN mens rea hierarchy as a sufficiently good culpability proxy to form the basis of a normatively justifiable criminal law.[19]

---

theorists, contemporary proponents include: GIDEON YAFFE, THE AGE OF CULPABILITY (2018); DOUGLAS HOUSAK, IGNORANCE OF THE LAW (2016); David Brink & Dana Nelkin, *Fairness and the Architecture of Responsibility*, *in* 1 OXFORD STUDIES IN AGENCY AND RESPONSIBILITY 284 (David Shoemaker ed., 2013); GARY WATSON, AGENCY AND ANSWERABILITY: SELECTED ESSAYS (2004); Larry Alexander, *Insufficient Concern: A Unified Conception of Criminal Culpability,* 88 CALIF. L. REV. 931 (2000); and Dan Kahan & Martha Nussbaum, *Two Conceptions of Emotions in Criminal Law*, 96 COLUM. L. REV. 269 (1996).

18. *See, e.g.*, SCANLON, *supra* note 17, at 161-66.

19. *See, e.g.*, Kenneth W. Simons, *Rethinking Mental States,* 72 B.U. L. REV. 463, 490 (1992) ("The reigning hierarchy often works fairly well in translating underlying normative approaches [to] blameworthiness . . . into doctrinal requirements."); Douglas Husak, *"Broad" Culpability and the Retributivist Dream*, 9 OHIO ST. J. CRIM. L. 449, 454-55 (2012) ("Ceteris paribus, a defendant who performs the actus reus of a crime purposely is more blameworthy than one who acts

In contrast, Part II demonstrates that the reasons-responsiveness picture of subjective culpability leads not just to intrahierarchical differences—cases where the reasons-responsiveness picture attributes differing degrees of culpability to two similarly situated offenders with the same PKRN mens rea states—but to *inter*hierarchical differences in culpability attribution as well. That is, for the same offense, the reasons-responsiveness picture will sometimes attribute more culpability to certain agents with a lesser mens rea state on the PKRN hierarchy, such as recklessness or negligence, than to someone with a higher mens rea state on the PKRN hierarchy, such as purpose. I argue that these cases of interhierarchical variance are not merely conceptually possible but are likely to be widespread in some of the most important cases of criminal homicide liability. I argue that we should expect cases of negligent or reckless criminal homicide—like that of Derek Chauvin's second-degree manslaughter conviction for negligently causing the death of George Floyd[20]—which involve the least criminal liability in the MPC mens rea regime, to frequently involve substantially more culpability than the typical purposeful homicide, which subjects agents to the most liability on the MPC mens rea regime.[21] This means that the shift from a reasons-responsiveness conception of culpability to the PKRN mens rea model involves not just a loss of information and so minor differences in the culpability of equally liable actors, but substantial and systematic mismatches between offenders' culpability and criminal liability. In addition to these important normative results, I also show how closer attention to these interhierarchical differences can provide an alternative, and compelling, explanation for recent empirical results concerning experimental subjects' sorting of traditional PKRN mens rea states that run counter to the MPC's expectations.

Finally, Part III considers how we might amend the MPC to accommodate a reasons-responsiveness picture of culpability, in light of the existence of interhierarchical variance. While defending a particular proposal is not the primary purpose of this Note, the Note does attempt to show how much of the previous discussion of reform presupposed that no significant interhierarchical variance was possible, and to explain how the costs and consequences of various reform options change once the prevalence of such variance is acknowledged. I survey several possibilities, including abandoning the requirement that criminal liability be weakly proportional to culpability, fundamentally revising the MPC's

knowingly, who in turn is more blameworthy than one who acts recklessly, who in turn is more blameworthy than one who acts negligently, who in turn is more blameworthy than one who is strictly liable because he acts with no culpability at all.").

20. State v. Chauvin, No. 27-CR-20-12646, 2021 WL 1559176 (Minn. Dist. Ct. Apr. 20, 2021) (verdict as to Count III), *appeal docketed*, No. A21-1228 (Minn. Ct. App. Sept. 23, 2021).

21. *See* MODEL PENAL CODE § 210.1-.4 (AM. L. INST. 1962).

mens rea regime, abandoning criminal grading in favor of increased judicial sentencing discretion, and introducing a general "absence of ill will" affirmative defense. Despite shortcomings with each solution, I argue that the "absence of ill will" affirmative defense is the least problematic of the possible solutions.

I. TENSIONS BETWEEN THE REASONS-RESPONSIVENESS THEORY AND THE PKRN MENS REA REGIME

A. The Reasons-Responsiveness Conception of Culpability

According to the reasons-responsiveness conception of culpability, an agent's culpability is a function of their reasons for acting in a certain way and the responses to those reasons which their actions reflect. These reasons are important for subjective culpability because it is the agent's reasons for acting that reveal (or constitute) their "quality of will." It is the agent's will, rather than their actions, which, according to the reasons-responsiveness theory, is the ultimate object of assessment, or grounds for our blame or resentment, of the culpable agent.[22] As philosopher Pamela Hieronymi puts the point:

> [T]he . . . activities, attitudes, and states of affairs for which [an agent is responsible] . . . will also reveal something of one's mind, of one's take on the world and what is important or worthwhile or valuable in it. By settling certain questions for oneself, by having a take on what is true, what is important, and what is to be done, one thereby constitutes those bits of one's mind relevant to the quality of one's relations with others—and so establishes what we might call one's *moral personality*, or, in an older but apt phrase, the *quality of one's will* . . . that bit of one's mind—one's moral personality or one's will (broadly construed)—just is the object of moral assessment and reaction. It is that which we assess, when we assess whether someone is morally praise- or blameworthy.[23]

The idea is that, if we look to when we hold agents responsible for their actions, and blame them for their wrongdoing, that blame appears to track not just whether some action was intentional or unintentional (as in the PKRN picture), but also more fine-grained motivational facts about what that intentional action illustrates about the agent's underlying values. In particular, our blaming practices appear to track what the action illustrates about how much the agent values, or fails to value, *us*—our life, well-being, or freedom. Blame tracks the agent's

---

22. *See, e.g.*, Kahan & Nussbaum, *supra* note 17, at 301-05.

23. Hieronymi, *supra* note 17, at 361-62.

responses to various reasons for and against intending to act, which help determine their quality of will—that is, their attitudes toward how important or unimportant, good or bad, valuable or not valuable, worthwhile or not worthwhile, some part of the world, such as our well-being, consent, or property rights, is.

As Peter Strawson, the progenitor of the modern reasons-responsiveness theories, puts the point, when an agent causes us injury, say by shoving us, what matters to us in determining how much we resent them is not primarily the degree of harm we suffered, but rather what that action "reflect[s]" about how much they value us.[24] Though "the pain may be no less acute," it will matter to our view of their culpability, says Strawson, whether the shover was (1) "trying to help me" by pushing me out of harm's way, and their reason for shoving me was that it would prevent further pain; (2) acting out of a "malevolent wish to injure me," and their reason for shoving was that it would cause me further pain; (3) shoving me "in contemptuous disregard of my existence," and their reason for shoving was not to cause me pain but simply to get me out of the way as an obstacle to their continuing on their prior path; or (4) shoving me "accidentally" while intending to, for example, signal a taxi, and so never considering my pain at all.[25]

In each of these four cases, the agent's intentional act reflects a different quality of will toward the person being shoved. In (1), the agent took the other person's well-being as a positive reason to act, and so their action does not reveal the kind of ill will that it otherwise might have. Instead, it reflects the agent's good will toward the other person. In (2) and (3), the agent's shoving shows that the agent failed to value the other person's well-being as much as they should have, but reflects varying degrees of ill will (or absence of appropriately strong good will). In (3) the agent's action reflects a failure to appropriately weigh the positive value of the other person's well-being as a reason to refrain from pushing them against the weight of their other goals, whereas in (2), the agent's action reflects an even deeper failure to appropriately weigh the value of the other person's well-being, failing to recognize their pain as a reason for refraining at all (indeed, treating their pain as a positive reason for acting). In (4), because the pushing was accidental, their action does not reflect a failure to appropriately weigh the value of the other person's well-being at all. At worst, it reflects minor ill will in the agent's negligent failure to respond to the reason that the other person's well-being provides them to take care to inquire into the possibility of accident.

---

24. Peter Strawson, *Freedom and Resentment*, 48 PROC. BRIT. ACAD. 1 (1962), *reprinted in* PERSPECTIVES ON MORAL RESPONSIBILITY 45, 48-50 (John M. Fisher & Mark Ravizza eds., 1993).

25. *Id.* at 48-49.

These differences in how the agent responded to the reason-giving force of another person's well-being when acting tracks, according to the reasons-responsiveness theory, the agent's culpability, or moral responsibility. Their degree of culpability is a function of the degree to which, in acting wrongfully, their actions showed a failure to value or appreciate the reasons why the criminal conduct they engaged in was wrongful. Criminal homicide is wrong, for instance, because of the value of the human life taken. The agent who murders another person is thus culpable because their murdering another person reveals that they failed to recognize the other person's life as valuable in the way that the law, by criminalizing homicide, insists they must.

B. Intrahierarchical Variance

It may not initially be clear just how distinct the two competing PKRN and reasons-responsiveness conceptions of culpability are or what practical difference those distinctions make. There is, of course, much in common between the reasons-responsiveness picture and the PKRN picture. The PKRN picture, like the reasons-responsiveness picture, begins with the insight that it is not just the action, but also the intention behind the action, that matters. As Oliver Wendell Holmes, Jr., put the point in *The Common Law*, "even a dog distinguishes between being stumbled over and being kicked."[26] The PKRN mens rea regime is designed, in part, to codify that basic insight, just as Strawson's discussion of quality of will is meant to explain our differing reactions to the four variants of the shoving agent from the previous Section.

Indeed, Strawson's variants may appear to track neatly onto the PKRN regime. The agent who purposefully pushes me (agent 2) shows more ill will toward me than the agent who merely knowingly pushes me (agent 3) on the way toward achieving some other purpose, who still shows more ill will toward me than the agent who accidentally pushes me through mere recklessness or negligence (agent 4). Still, while the precise limits that this mapping places on the degree of variance between the reasons-responsiveness and PKRN conceptions of subjective culpability is the central subject of Part II, it is clear that at least some variance is possible.

One initial way of putting the theoretical difference between the two pictures is in their psychological object. According to the reasons-responsiveness theory, what matters in assigning blame is not just whether the action was intentional, but the reasons or motivations behind the intention. In the psychological path toward an agent's action, the two pictures differ in how far back along the path of the agent's subjective psychology they locate the agent's culpable mens rea.

---

26. OLIVER WENDELL HOLMES, JR., THE COMMON LAW 5 (Harvard Univ. Press 2009) (1881).

Take, for example, the case of murder, involving an agent who deliberates and decides to kill another person, forms an intention to kill them, and then intentionally causes their death. We can trace the volitional story as follows:

> Stage 1: reasoning about whether to cause the death of another person;
> Stage 2: forming an intention to cause the death of another person;
> Stage 3: causing the death of another person.

Whereas the PKRN regime concentrates on the agent's intention (stage 2) as the locus of their culpable mens rea, the reasons-responsiveness account concentrates on a more distal mental state — the reasons, or reasoning process, that produced the intention (stage 1).

Lest this seem like a distinction without a difference, this move further down the causal chain of action gives the reasons-responsiveness account the conceptual tools to distinguish between the degree of culpability in various intentions to cause the death of another, just like moving down the chain from action to intention lets the PKRN model distinguish between the culpability of intentional and nonintentional performances of the same action. Just as, for the PKRN picture, various agents who all cause the death of another person in stage 3 might be differently culpable depending on whether they were purposeful, knowledgeable, reckless, or negligent during stage 2, on the reasons-responsiveness account, various agents who all intended to cause the death of another in stage 2 and so caused the death of another might be differently culpable depending on the reasons for which they formed the intention to cause the death during stage 1.[27]

To see the difference in action, consider *People v. Kevorkian*, where Dr. Jack Kevorkian was charged with first-degree murder for aiding terminally ill patients with extreme suffering in assisted suicide.[28] Under the PKRN picture of culpability, it does not matter what reason Kevorkian had for intending to cause the death of another person. An intention to kill is the most atomic level of analysis for culpability. As such, the PKRN model lacks the resources to treat Kevorkian's actions, if they amount to an intentional killing, any differently than any other intentional killing. On the reasons-responsiveness picture, however, what matters for Kevorkian's subjective culpability is not *whether* Kevorkian intended to kill his terminally ill patients, but *why*. What distinguishes Kevorkian's case from a typical intended causing of another's death is that a typical intention to

27. The intention can be *multiply realized* by a variety of more fine-grained sets of reasons-responses, each of which can cause, or constitute, the intention to act. For a more detailed discussion of multiple realizability in the philosophy of mind, see Hilary Putnam, *Psychological Predicates*, *in* ART, MIND, AND RELIGION 37, 37-48 (W.H. Capitan & D.D. Merrill eds., 1967).

28. 517 N.W.2d 293, 295 (Mich. Ct. App.), *vacated*, 527 N.W.2d 714 (Mich. 1994).

cause another's death treats some reason (inheritance, revenge, etc.) as more important than avoiding the suffering of the one who dies, whereas in Kevorkian's case, it was precisely his recognition of the importance of avoiding the suffering of the one who dies that formed his reason for intending the death in the first place. While the content of his intention was the same, his reasons for intending, and thus the quality of will he bore toward the one he killed, were radically different.[29]

While Dr. Kevorkian's case is an extreme, if illustrative, example, there will also be room on the reasons-responsiveness account for much more fine-grained culpability distinctions in more quotidian cases of criminal wrongdoing. In fact, there will be as many different degrees of culpability for the agent who intentionally acts as there are reasons for so intentionally acting. Just as, on the PKRN picture of subjective culpability, the same agent who performs the same action might be differently culpable depending on whether the action was intended, here the same agent who intentionally acted might be differently culpable depending on whether, and to what degree, the intended action demonstrated a faulty response to reasons or not.[30]

Consider another example, drawn from a study by Mark D. Alicke.[31] If I intentionally run a red light because I am late to pick up my child from school (a reason to which I appropriately assign a relatively high weight) and you intentionally run a red light because you are late for your favorite television episode rerun that you were planning to watch when you got home (a reason to which you appropriately assign a low weight), on the reasons-responsiveness picture, my running the red light is consistent with my being less culpable than you are for the same intentional action. The stronger my competing reason for running the light, the less I mistakenly undervalue the other drivers whose well-being I risk when speeding through an intersection. Importantly, my stronger reason

---

29. One need not even agree that Kevorkian was *correct* in his response to his reasons to think that he was at least less evil (on the reasons-responsiveness model) than the murderer who kills their victim for the inheritance or the sociopath who kills their victim out of boredom.

30. Indeed, the Purpose, Knowledge, Recklessness, and Negligence (PKRN) grading regime for criminal homicide under both the MPC and the non-MPC regimes that follow the alternate Penn System, *see supra* note 5, acknowledge these subtle variations within intentional killings or reckless killings, *see supra* notes 7-8 and accompanying text.

31. Mark D. Alicke, *Culpable Causation,* 63 J. PERSONALITY & SOC. PSYCH. 368, 369 (1992). While varying the countervailing reasons is one way of generating these kinds of cases of intrahierarchical culpability variance, it is not the only way. *See, e.g.*, LARRY ALEXANDER, KIMBERLY KESSLER FERZAN & STEPHEN J. MORSE, CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW 23-65 (2009) (creating similar counterexamples by changing the degree of subjective probability assigned to a risk by reckless agents); YAFFE, *supra* note 17, at 91-93 (creating similar counterexamples by changing reckless actors' counterfactual dispositions to act given varying degrees of subjective probability or risk).

does not justify, nor does it excuse, my behavior. If I were properly responding to reasons, I would recognize that my reasons to pick up my child on time do not outweigh the risk I pose to other drivers. But it can, on the reasons-responsiveness picture, make my behavior less culpable than if I had given no thought to the needs of the other drivers at all. While both agents will be deemed equally culpable on the PKRN model, they will not be equally culpable on the reasons-responsiveness model, because their purposeful actions, though identical, do not evince the same degree of ill will toward others. The same sorts of cases can be constructed for foreseen but unintended consequences by varying the quality of the defendant's reasons for acting in the way that would produce the foreseen harm, and for cases of recklessness by varying the quality of the defendant's reasons for acting in the way that would create the foreseen risk of that harm.

To be sure, unlike the purpose and knowledge mens rea elements, recklessness and negligence do incorporate into their definition some question of the strength of the defendant's subjective reasons for their action.[32] The MPC requires that the reckless agent's perceived risk of the material element be "unjustifiable" given their subjective reasons for acting, and that it constitute a "gross deviation" from the actions of a reasonable person.[33] However, though this "unjustifiability" standard for recklessness (like the affirmative defenses of necessity or self-defense) does provide a limited role for an agent's reasons in the MPC mens rea categories, it is still insufficient to provide the kinds of fine-grained culpability distinctions that the reasons-responsiveness conception entails among reckless agents. This is because while the assessment of reasons, and thus culpability, on the reasons-responsiveness picture involves gradations in the quality of reasons, the MPC unjustifiability condition is a threshold measure. If one's reasons are sufficiently strong, one is not reckless. If one's reasons fall below that standard and so are unjustifiable, the agent is reckless, regardless of how close to — or far from — the standard the agent's reasons were.

This means that a reasons-responsiveness conception of culpability will still entail differences in culpability among reckless agents whose reasons fall below the justifiability threshold when the strength of their countervailing reasons falls short to varying degrees. Consider our previous example where two reckless

32. *See* MODEL PENAL CODE § 2.02(2)(c) (AM. L. INST. 1962) ("A person acts recklessly with respect to a material element of an offense when he consciously disregards a substantial *and unjustifiable* risk that the material element exists or will result from his conduct. The risk must be of such a nature and degree that, considering the nature and purpose of the actor's conduct and the circumstances known to him, *its disregard involves a gross deviation from the standard of conduct that a law-abiding person would observe in the actor's situation*." (emphases added)). Similar qualifiers exist for the MPC's definition of negligence. *See id.* § 2.02(2)(d).

33. *Id.* § 2.02(2)(c).

agents run a red light. Suppose both drivers believe that there is a seventy percent chance that they will run a red light while speeding through an intersection but do so anyway: one in order to pick up their daughter, the other to watch their favorite television show. As neither reason is sufficient to justify risking running the red light, both act in ways that are "unjustifiable" given their reasons for acting—and in ways that constitute a "gross violation" of the reasonable-person standard. Both drivers would thus count as reckless under the MPC, and so be equally culpable on the PKRN conception of culpability. Nonetheless, as we have seen, the two reckless agents would be differently culpable according to the reasons-responsiveness picture. Although both actors engaged in "unjustifiable" actions and were therefore culpable, we can still consider the question of *how* unjustifiable their actions were. If, given the respective weight of their reasons for acting, the first agent behaved in a way that was less unjustifiable than the second agent, the reasons-responsiveness conception of culpability would hold these two actors to be culpable to different degrees.[34]

Though some theorists have questioned the appropriateness of negligence liability on a reasons-responsiveness account,[35] there seems to be no reason in principle why a theorist committed to the reasons-responsiveness conception of culpability cannot distinguish between culpability in negligence cases in the same way. Suppose two ship captains, *C1* and *C2*, negligently fail to fully inquire about the seaworthiness of their vessel, revealing in their failure of attention and inquiry an insufficient regard for the safety of their passengers. Suppose *C1* has some stronger reason than *C2* for hurrying that day. Let us suppose, as in the case above, that *C1* wants to be on time to pick up their child and that *C2* wants to be on time for their nightly television viewing. As in the case of intentional action, here the stronger the weight of the negligent defendant's reason for their failure to attend, the less they must culpably devalue the reasons provided by the well-being of their passengers, which count in favor of inquiring about the possibility of a risk to those passengers' well-being when leaving the dock. While

34. The same threshold formulation of unjustifiability allows for similar variance in degrees of "unjustifiable" failures to attend to risks among equally negligent actors. *See id.* § 2.02(2)(d).

35. *See, e.g.*, ALEXANDER ET AL., *supra* note 31, at 69-71. A dialectical point: while I will assume that negligent actors can evince failures of reasons-responsiveness in their failures to attend and inquire, nothing hangs on this assumption. Instead, I will argue in this Note that insofar as the MPC can successfully give an account of how negligent agents are culpable on a reasons-responsiveness picture, that account will entail that negligent agents can be *more culpable* than purposeful agents for the same offense, and so be incompatible with typical proportionality requirements for a justified system of criminal law. Readers who deny the reasons-responsiveness account of negligent agents' culpability suggested above will already accept the Note's conclusion that the standard attempts to reconcile reasons-responsiveness with the MPC are incapable of justifying the MPC's use of negligence liability.

both agents will be deemed equally culpable under the PKRN model,[36] they will not be equally culpable on the reasons-responsiveness model because their negligent actions, though identical, do not evince the same degree of ill will toward others.

C. The Normative Challenge of Intrahierarchical Variance

As we have seen, even though there may be similar motivations underlying the PRKN and reasons-responsiveness pictures of subjective culpability, the two are distinct in important ways. There are both theoretical differences in the ultimate object of subjective culpability—for PKRN, the proximate intentions of the agent for their actions, and for reasons-responsiveness, the more distal reasoning behind those proximate intentions—and extensional differences in the subjective culpability of wrongdoers within a PKRN grade. On the reasons-responsiveness picture there will be more and less culpable reckless actors, or purposeful actors, depending on the quality of their reasons for acting recklessly or purposefully.

This variance in culpability for equally liable defendants poses a potential difficulty for those criminal-law theorists who accept the reasons-responsiveness picture of criminal culpability while also accepting some version of what we might call the "weak proportionality principle": a minimally acceptable criminal system will not hold substantially less culpable agents substantially more criminally liable than substantially more culpable agents for the same criminal act. While most criminal-law theorists would not think that weak proportionality is sufficient to justify a system of criminal law, I take this requirement to be a necessary minimal standard of normative acceptability for any theory of criminal liability to meet.

This normative requirement of weak proportionality is weak in at least two important ways. First, it does not require that criminal liability be perfectly proportional to culpability. The American criminal grading systems have never purported to mark every difference in culpability, and Eighth Amendment jurisprudence makes clear that the requirements of proportionality condemn only punishments that are "greatly disproportioned" to the culpability involved in the offense charged.[37] Second, weak proportionality is weak in the sense that it requires no strong assumptions or commitments concerning any underlying view about the justification for the state's use of coercive force, the justification of

36. Assuming, again, that neither of the defendant's reasons suffice to prevent the negligence from constituting a "gross deviation" under the MPC. MODEL PENAL CODE § 2.02(2)(c) (AM. L. INST. 1962).

37. *See, e.g.*, O'Neil v. Vermont, 144 U.S. 323, 340 (1892).

punishment, or the function of criminal law.[38] Weak proportionality lays out a minimum requirement compatible with whatever mixture of mainstream expressivist, retributivist, reparative, rehabilitative, or deterrence-based grounds that criminal-law theorists find most appealing. Nonetheless, weak proportionality does create certain minimum requirements of alignment for a theory of criminal law between the theory's descriptive system for assigning criminal liability and its underlying normative picture of culpability or moral responsibility.

The above examples pose at least some challenge to the weak proportionality requirement: the fact that the mental states involved in reasons-responsiveness, which form the basis of criminal culpability, differ from the mental states of the PKRN hierarchy, which form the basis of the MPC's theory of criminal liability, opens up the possibility of misalignment between an agent's culpability and their criminal liability, creating pressure on the commitment to proportionality between the two. How much of a problem this is, however, depends on how much variance is possible between the two pictures. In Parts II and III, I will argue that the possible variance is greater than has previously been assumed, and so the normative problems posed by a joint commitment to the two is greater than typically thought.

If we limit ourselves to the examples of variance we have seen so far, however, it is not obvious that the failures of proportionality are sufficiently substantial to violate the weak proportionality requirement. If we canvas the cases raised in the prior discussion of the differences between the verdicts of the reasons-responsiveness picture and the PKRN picture concerning criminal culpability, one striking feature of each of them is that while differences exist in relative culpability between intentional or reckless agents, the weak ordinal culpability verdicts between the purposeful, reckless, knowledgeable, or negligent agent all appear to map on to the ordinal culpability verdicts of those agents in terms of their respective quality of will or response to reasons.

The differences in quality of will discussed in this Part all involve intrahierarchical differences in subjective culpability: cases where one purposeful homicide, for example, exhibits more ill will than another, or cases where one reckless agent values the well-being of the injured party less than another reckless agent. This is consistent with the idea that the quality of will of an offender with a higher grade of mens rea in the PKRN hierarchy is always (or at least typically) acting with more ill will than an agent with a lower grade of mens rea. That is, all those who intentionally kill another will typically show less value for human life than all those who are merely reckless or negligent toward another's death.

---

38. *See, e.g.*, RONALD DWORKIN, LAW'S EMPIRE 109-10 (1986) (describing the primary purpose of law as the justification of state coercion).

Indeed, reflection on Strawson's four actors may suggest a principled argument for concluding that the variance between culpability on both a PKRN mens rea regime and a reasons-responsiveness picture of criminal liability *must* be limited to intrahierarchical variance.[39] Consider a battery statute that prohibits the shoving of the Strawsonian agents, graded by mens rea. According to the reasons-responsiveness conception of culpability, statutes that criminalize such behavior require that agents take the fact that they are harming another as a reason to refrain from so acting.[40] Purposeful harm involves harming another as a reason that counts in favor of acting, rather than against acting. This is a bigger failure from the ideal reasons-responsiveness perspective than is merely failing to take the harm as a sufficiently compelling reason to refrain from an action that foreseeably causes that harm. Recklessly acting with only a risk of harm will involve yet less undervaluing of the reasons the victim's well-being provides to refrain, since the reckless actor, unlike the knowing actor, can discount that harm by the believed likelihood the harm will result. Failures of attention associated with negligence will involve even more minor deviation from the proper response to the positive reason the victim's well-being provides to positively engage in seeking more information about one's actions when one does not yet realize the risk involved.

While certain reckless agents may exhibit a lower degree of ill will toward their victims than other reckless agents, depending on the degree of probability they assigned to their actions' harming the victim, they will necessarily have had less ill will than they would have had they been similarly situated but acted knowing for certain that they would cause harm to their victim.[41] And while various knowing agents might, depending on their reasons for pursuing their goals, exhibit varying degrees of ill will in accepting the foreseen harms to the victim, surely they exhibit *more* ill will when those harms are part of their goal. Call this the "argument for ordinal convergence." If it were sound, this argument would provide theoretical grounds for confidence that the kinds of examples surveyed above, which are limited to intrahierarchical variance, are the only kinds of variance that could possibly be discovered. If one could provide a satisfactory account of such variance, one would have thus succeeded in reconciling the tensions between the MPC and the reasons-responsiveness conception of culpability.

---

39. *See supra* notes 24-26 and accompanying text.

40. *See, e.g.*, YAFFE, *supra* note 17, at 70-71.

41. *See* ALEXANDER ET AL., *supra* note 31, at 62-95.

D. Reconciling Intrahierarchical Variance: PKRN as Culpability Proxies

Suppose for the moment that this claim about the limit to intrahierarchical differences is correct and that PKRN are successful ordinal culpability proxies, such that an agent who intends to murder always values life less than the reckless agent, even if certain reckless agents value life more or less than other reckless agents and certain intentional murderers value life more or less than other intentional murderers. Is that sufficient to show that using PKRN proxies to model the real sources of subjective culpability—an agent's underlying reasoning—is permissible? A strong case can be made in the affirmative. It is no mark against the PKRN grading system that it groups together all the reckless agents, or purposeful agents, or negligent agents, even if those agents all exhibit varying degrees of culpability. In fact, there may be important reasons to do so. Moreover, a closer look at the MPC shows that a reading of PKRN as proxies for quality of will as an underlying source of culpability may actually provide a better account of the MPC than would a picture of the MPC as grounded entirely in a PKRN picture of subjective culpability.

Consider first the picture of proportionality underlying a normatively justifiable criminal law. Proportionality may require that a criminal grading system not treat certain less culpable agents as more criminally liable than more culpable agents. But the criminal law has never purported to mark every difference in culpability. As we have seen, American law seems to adopt a picture of weak proportionality according to which a just criminal-law system will not hold more liable any less culpable criminal offender than it does a more culpable criminal offender for the same act. And intrahierarchical differences in subjective culpability among the PKRN categories is consistent with such a requirement. If a mens rea regime treats all reckless agents the same, it will fail to punish more some reckless agents who are more culpable, but it also will avoid punishing more any reckless agents who are less culpable.

Second, we can note that proportionality does not require that we identify in a criminal statute the actual source of an agent's culpability, so long as the feature of the agent's subjective psychology that we do identify follows what Doug Husak has referred to as the "equal culpability thesis": the requirement that an agent who is in that substituted mental state is as culpable (or more) than would be the agent with the mental state actually required by the statute.[42] Evidence of the equal-culpability thesis is commonplace in the American legal system, where the criminal law regularly permits mens rea substitution principles, allowing the

---

42. *See, e.g.*, Douglas N. Husak & Craig A. Callender, *Willful Ignorance, Knowledge, and the "Equal Culpability" Thesis: A Study of the Deeper Significance of the Principle of Legality*, 1994 WIS. L. REV. 29, 53-58.

state to prove some mental state other than the statutorily required mens rea, provided that that substituted mens rea is as bad, or worse, than the mens rea being substituted for. The acceptance of such principles is perhaps most obvious in the MPC's discussion of PKRN, where it allows that establishing the existence of some mental state higher in the PKRN hierarchy can substitute for whatever PKRN mental state is actually required by a criminal statute.[43]

We can see a similar principle at work in the old common-law murder doctrine of transferred intent, whereby an intention to kill person *A* can substitute for the intention to kill some bystander *B*, on the theory that intending to kill *A* is just as culpable a mental state as intending to kill *B*.[44] More recently, mens rea substitution principles can be seen in the federal courts' increasingly well-established "willful-blindness doctrine," which treats willful ignorance as a substitute for knowledge as a mens rea element,[45] and is understood as being justified by many on the principle that willful ignorance is equally culpable.[46] The use of such proxies is even more apparent in the material elements of "proxy crimes," where some nonharmful or less harmful material element (e.g., failure to comply with environmental reporting and record-keeping) is included as a proxy for a

43. MODEL PENAL CODE § 2.02(5) (AM. L. INST. 1962).

44. *See* Nancy Ehrenreich, *Attempt, Merger, and Transferred Intent*, 82 BROOK. L. REV. 49, 51 (2016); *see also* People v. Scott, 14 Cal. 4th 544, 545 (1996) ("Under the classic formulation of California's common law doctrine of transferred intent, a defendant who shoots with the intent to kill a certain person and hits a bystander instead is subject to the same criminal liability that would have been imposed had 'the fatal blow reached the person for whom it was intended.' *In such a factual setting, the defendant is deemed as culpable as if he had accomplished what he set out to do*." (emphasis added) (quoting People v. Suesser, 142 Cal. 354, 366 (1904))).

45. *See* Global-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 766 (2011) ("The traditional rationale for this doctrine is that defendants who behave in this manner are just as culpable as those who have actual knowledge."); *accord* United States v. Heredia, 483 F.3d 913, 917, 924 (9th Cir. 2007) (en banc); United States v. Jewell, 532 F.2d 697, 702-03 (9th Cir. 1976) (en banc).

46. *See*, *e.g.*, Husak & Callender, *supra* note 42, at 35-36; Gideon Yaffe, *The Point of Mens Rea: The Case of Willful Ignorance*, 12 CRIM. L. & PHIL. 19, 19 (2018) (justifying the place of willful ignorance, and the PKRN mens rea hierarchy more generally, through its function as a proxy for the agent's underlying reasons-responsiveness); Alexander F. Sarch, *Willful Ignorance, Culpability and the Criminal Law*, 88 ST. JOHN'S L. REV. 1023, 1026-27 (2014). *But see Global-Tech Appliances*, 563 U.S. at 773 (Kennedy, J., dissenting) (arguing against this "traditional rationale" of the willful-ignorance doctrine).

correlated, more serious, but harder to prove, wrong-making element (e.g., engaging in unreported discharge of pollutants).[47] Proportionality might even require such proxies, given the difficulty in determining an agent's reasons and the danger that attempts to make more fine-grained distinctions would introduce more mistakes in culpability attributions than they would fix.[48]

Finally, the law has other mechanisms besides its articulation of the elements of the prima facie case that it can employ in efforts to fine-tune its apportionment of criminal liability to accommodate subtle differences in culpability. Differences in culpability between equally criminally liable offenders can be taken into account at the sentencing stage,[49] as well as by various affirmative defenses, which can function as complete, or partial, shields to criminal liability.[50] In fact, assuming only intrahierarchical variance, not only do these various points show that the MPC's PKRN grading scheme could be consistent with an underlying reasons-responsiveness picture of subjective culpability, as well as with an underlying PKRN picture of subjective culpability, these points also help suggest that perhaps the MPC's criminal liability regime actually does presuppose an underlying reasons-responsiveness picture of subjective culpability. If we look for a common thread in affirmative defenses like imperfect self-defense or duress in sentencing guidelines, or in mens rea substitution principles like willful ignorance, many of them appear designed precisely to track subtle differences in the reasons of various agents within a class of purposeful, knowledgeable, reckless, or negligent agents.

Accordingly, we might read the MPC as being designed for a picture of subjective culpability where an agent's quality of will is their ultimate source of culpability, and where PKRN mens rea elements will involve only intrahierarchical

---

47. *See, e.g.*, 42 U.S.C. § 6928(d)(4) (2018) (criminalizing such failures to report under the Resource Conservation and Recovery Act). In addition to reporting crimes, criminal-law theorists have understood a variety of possession crimes, such as possession of drugs or child pornography, as similar proxy crimes. *See, e.g.*, Anthony M. Dillof, *Possession, Child Pornography, and Proportionality: Criminal Liability for Aggregate Harm Offenses*, 44 FLA. ST. U. L. REV. 1331, 1333 (2017); Douglas Husak, *Drug Proscriptions as Proxy Crimes*, 36 LAW & PHIL. 345, 348-39 (2017); Melissa Hamilton, *The Efficacy of Severe Child Pornography Sentencing: Empirical Validity or Political Rhetoric?*, 22 STAN. L. & POL'Y REV. 545, 548-49, 560-61 (2011).

48. *See, e.g.*, Holly Lawford-Smith, Book Review, 35 AUSTRALIAN J. LEGAL PHIL. 152, 157-58 (2010) (reviewing ALEXANDER ET AL., *supra* note 31) (noting the "utter impractica[bility]" of using reasons-responsiveness to replace PKRN as the mens rea element in criminal law).

49. *See* U.S. SENT'G GUIDELINES MANUAL ch. 3 (U.S. SENT'G COMM'N 2004) (detailing adjustments to be made to a criminal defendant's sentence based on certain motivational factors). The view that certain culpable reasons that society deems especially problematic ought to subject an agent to more criminal liability has been growing in popularity, both for sentencing and for criminal liability in special cases, such as "hate crimes." *See* Carissa Byrne Hessick, *Motive's Role in Criminal Punishment*, 80 S. CAL. L. REV. 89, 89 (2006).

50. *See supra* note 9 and accompanying text.

differences in culpable quality of will. Such a picture can account not just for the PKRN grading as culpability proxies, but also for the places where the MPC varies from the general PKRN grading schema. Those variances are attempts to build into the model the ability to make more fine-grained culpability distinctions within grades at different places within the criminal penal code that would be otherwise inexplicable on the PKRN picture of subjective culpability.

II. INTERHIERARCHICAL VARIANCE BETWEEN THE REASONS-RESPONSIVENESS AND PKRN ACCOUNTS OF SUBJECTIVE CULPABILITY

A. Showing Interhierarchical Variance

We have seen that despite the apparent reliance of the MPC's PKRN grading regime on the PKRN picture of subjective culpability, it is possible to fit the MPC into a reasons-responsiveness picture of subjective culpability, even if the reasons-responsiveness picture of subjective culpability gives different verdicts about culpability from the PKRN picture of subjective culpability, so long as those differences are merely intrahierarchical.

As I argue in this Part, however, the problem with such a fitting is that the differences are *not* merely intrahierarchical. Theorists have failed to notice that the same sorts of differences that help generate *intra*hierarchical differences between the reasons-responsiveness and PKRN picture will also consistently generate *inter*hierarchical differences. That is, cases where for the same act, on the reasons-responsiveness picture a purposeful agent is less culpable than a reckless agent, or a knowledgeable agent is less culpable than a negligent agent. Indeed, as we will see, for certain crimes on the reasons-responsiveness picture, we should *expect* that purposeful agents will be systematically less culpable than reckless agents. The MPC's PKRN grading schema is thus not merely a harmless (and perhaps helpful) abstraction for modeling subjective culpability. In fact, it threatens to introduce widespread errors and disproportionate punishment on the reasons-responsiveness picture.

1. A General Recipe for Constructing Examples of Interhierarchical Differences

The central problem with the conclusion that the tensions between the MPC and a reasons-responsiveness picture of culpability are exclusively intrahierarchical is that different qualities of will can vary not just within PKRN mental-state categories but also *across* categories. Recall the method by which we generated our cases of intrahierarchical differences, by holding fixed the purposeful

wrongful action (or knowing, reckless, or negligent action) and varying the strength of the countervailing reasons that, while all still insufficient to avoid criminal culpability, were motivating the agent to act. The stronger those countervailing reasons, the more the agent might have valued the well-being of the wronged victim, and so the less culpable ill will they would have demonstrated.

To construct cases of interhierarchical variance, we can modify the cases of intrahierarchical variance from Part I. Imagine two drivers, *D1* and *D2*, each of whom are rushing home while thinking there is a fifty percent chance they are going over the speed limit, and so act recklessly. *D1* has no strong reason to be rushing, but simply does not care about the risk that they will pose to other drivers if they drive above the speed limit. Given the low subjective value they place on their reason for driving quickly, they must, even with a fifty percent discount, place a very low value on obeying the speed limit to have acted as they did. In contrast, *D2* has a much stronger (though still insufficient to avoid the "unjustifiability" threshold) reason to be rushing home[51]: they are late to relieve their child's babysitter. Suppose that *D2*'s subjective reason is more than twice as strong as *D1*'s. While their reason for getting home does not justify driving above the speed limit and endangering other drivers (and would not rise to the level of qualifying for an affirmative defense like duress or necessity), it comes close to putting them outside the realm of gross deviation. *D2* need not be indifferent to the reason-giving force of the harm they might cause to other drivers. Because of the strength of their positive reasons for getting home, *D2* might treat the causing of harm to others as a serious reason to refrain from acting, yet still risk speeding, due to the strength of those countervailing reasons.

Now imagine a third driver, *D3*, who purposefully exceeds the speed limit rushing home for the same reason as *D2*: to relieve their child's babysitter. Because their reason for rushing is stronger than *D1*'s, their purposefully exceeding the speed limit is more consistent with caring about the well-being of other drivers. *D3*'s reason for purposefully exceeding the speed limit is more than twice as strong as *D1*'s reason, and so consistent with less ill will toward other drivers, even when driving above the speed limit was intentional. The purposeful agent with a stronger reason for acting may commit an offense purposefully in spite of their granting more weight in their reasoning to the harm that they will cause through that offense than does the reckless agent with weaker reasons for acting, who must be indifferent to the harm they cause in committing the offense, even if they are only weighing their actions against a risk of committing that offense.

Here we have three not atypical cases:

Case 1: Weak-Reason Reckless Action

---

51. *See supra* notes 32-34 and accompanying text.

Case 2: Strong-Reason Reckless Action
Case 3: Strong-Reason Purposeful Action

According to the reasons-responsiveness theorist, the order of culpability from most culpable to least culpable is (1), (3), (2). According to the PKRN picture of subjective culpability (and thus the PKRN grading system) the order is (3), (1 & 2). If we follow the PKRN grading system, we will not simply be losing information about the degree of culpability across offenders; we will get different orderings, and thus disproportionate assignments of criminal liability.[52]

2. *The Normative Significance of Interhierarchical Culpability Differences*

How big of a problem is the existence of such interhierarchical differences? While the introduction of errors is worse than the mere loss of information, the severity of the threat depends on how widespread and how serious the errors are. In fact, if the reasons-responsiveness picture is correct, then there is no reason to think errors are not widespread, and some reason, in especially important cases like criminal homicide, to be almost certain that they are. There will be more than occasional interhierarchical variance in assignments of culpability for various criminal wrongdoing on the two pictures. The differences will be as widespread as there are variances in strengths that a defendant places on the weight of others' well-being as a reason to refrain from wrongdoing and the weight of the relative countervailing reasons to so act.

Consider again a variant of our previous example. *D1* and *D2* are each speeding home. *D1* knows that she is going over the speed limit, whereas *D2* merely believes there is a fifty percent chance that she is doing so and is thus reckless with regard to speeding. Suppose further that *D1* is speeding because she is late to pick up her child (which she accurately values at some value 2*X*) and *D2* is speeding home to watch her favorite evening television episode (which she accurately values at *X*). In this case, assuming some standard expected utility calculus, it would take precisely the same weighing of the force of the well-being of other drivers, (*X*-1), to explain their actions.[53] As such, it will be purely a matter

52. One may object that this case relies upon the use of a circumstance element under MODEL PENAL CODE § 1.13(9)(ii) (AM. L. INST. 1962), where the differences between purposeful agents and knowledgeable agents tend to blur. One may thus worry that I have not (yet) given a case where a purposeful agent (rather than merely a knowledgeable agent) is more culpable than a reckless agent for the same result or act element. In Section II.B, *infra*, I will show in more detail how a purposeful agent may be less culpable than a negligent or reckless agent, even for their purposeful conduct or for results they purposefully bring about.

53. This factors into *D2*'s discounting of the value they place on not harming others through speeding by their expected probability that they will produce such a harm.

of chance which of these two drivers happens to be more culpable, depending on which one happened to weigh less the value of the well-being of the other drivers. In such circumstances, we should expect the PKRN grading system to give wrong results as often as not. So long as the various reasons *D1* and *D2* have in favor of wrongdoing differ in strength to a sufficient degree, these differences will be enough to swamp differences between the subjective value they place on the well-being of others, which determines their subjective culpability on the reasons-responsiveness picture.[54]

3. Examples of Interhierarchical Differences in Criminal Homicide Case Law

It gets worse. The method of designing cases of interhierarchical variance described in the previous Sections assumes that there is a random distribution of reasons for acting in the commission of various crimes across purposeful, knowing, reckless, or negligent actors. For some crimes, this is unlikely. Most strikingly, criminal homicide seems like a troubling example where we can predict systematic mismatches where the negligent or reckless agent will have a *worse* quality of will than a purposeful agent.

Consider first the purposeful agent. While there are occasional sociopaths,[55] a typical purposeful killing involves more intelligible motives. While such motives obviously do not excuse murder, they also do not evince the same lack of value for human life as someone, like the mobster hitman, who is willing to take the life of another person for minor financial or personal gain.

In contrast, someone who recklessly risks taking a life may be more likely to do so following a cold calculation concerning expected financial gain. The landlord who knows there is a five percent chance that their failure to upgrade their building's fire alarm system will kill at least one of their tenants but fails to repair it because of the cost is far closer to the mobster hitman than is the typical purposeful killer with more intelligible motives.

To make this theoretical point more concrete, consider the cases of negligent homicide in *Commonwealth v. Welansky*,[56] which involved a nightclub owner who neglected to inquire into the status of the building's emergency exits, resulting

54. This is not the only possible method to construct cases of interhierarchical difference. Similar cases could be constructed by, for example, varying the counterfactual dispositions of the two agents to create reckless agents who are more culpable than knowing or purposeful agents.

55. See, for example, the case of Robert Harris, described in Gary Watson, *Responsibility and the Limits of Evil*: *Variations on a Strawsonian Theme*, *in* PERSPECTIVES ON MORAL RESPONSIBILITY, *supra* note 24, at 119, 131-43.

56. 55 N.E.2d 902 (Mass. 1944).

in the death of his patrons when the building caught fire and they could not escape, and *State v. Chauvin*,[57] where police officer Derek Chauvin neglected to consider the possibility that his kneeling on George Floyd's neck for nine minutes and twenty-nine seconds could result in his death.[58]

Compare these cases with the study of female-perpetrated purposeful homicides from Angela Browne and Kirk R. Williams, who note that—particularly in the context of partner homicides—somewhere between seventy-five and ninety-three percent of female perpetrators report having been physically assaulted or abused by the victim prior to the murder.[59] Even for those cases of past abuse which may not meet the criteria of available affirmative defenses such as that of Battered Person Syndrome or self-defense,[60] the resulting purposeful homicides are still often a result of an absence of "legal [or] extra[-]legal resources"—such as the historical absence of non-life-threatening abuse as grounds for divorce action, failures of past police interventions, or the absence of available shelters[61]—that can lead to an "overwhelming and entrapping life situation" that motivates the agent to purposefully cause the death of another person.[62] While these intelligible motives need not negate their criminal culpability on the reasons-responsiveness picture, it will make them less culpable than

57. No. 27-CR-20-12646, 2021 WL 1559176 (Minn. Dist. Ct. Apr. 20, 2021) (verdict as to Count III), *appeal docketed*, No. A21-1228 (Minn. Ct. App. Sept. 23, 2021).

58. In both cases, the defendants were ultimately convicted not merely of negligence but of recklessness (and, in Chauvin's case, the strict-liability crime of felony homicide). *See Welansky*, 55 N.E.2d at 908, 913; *Chauvin*, No. 27-CR-20-12646, 2021 WL 1559182 (verdict as to Count I), *appeal docketed*, No. A21-1228 (Minn. Ct. App. Sept. 23, 2021); *Chauvin*, No. 27-CR-20-12646, 2021 WL 1559174 (verdict as to Count II), *appeal docketed*, No. A21-1228 (Minn. Ct. App. Sept. 23, 2021); *Chauvin*, No. 27-CR-20-12646, 2021 WL 1559176 (verdict as to Count III), *appeal docketed*, No. A21-1228 (Minn. Ct. App. Sept. 23, 2021). However, it is not clear (especially in *Welansky*) whether these verdicts reflected an accurate application of the mens rea elements as articulated in the PKRN regime, or whether they reflected normative pressure on the juries and judges, based on the defendants' obvious culpability, to hold those defendants more criminally liable in ways that the PKRN mens rea regime cannot actually accommodate. Furthermore, whether the crime was reckless or negligent does not seem to explain the culpability involved. Supposing Chauvin or Welansky had actually been only negligent, and not reckless, many would still have the strong intuition that such crimes were examples of a high degree of culpability, not just equal to, but exceeding, the culpability in the typical purposeful homicide.

59. Angela Browne & Kirk R. Williams, *Exploring the Effect of Resource Availability and the Likelihood of Female-Perpetrated Homicides,* 23 LAW & SOC'Y REV. 75, 77 (1989).

60. *Id.* at 75-80.

61. *Id.* at 78-79.

62. *Id.* at 80 n.6 (quoting JANE TOTMAN, THE MURDERESS: A PSYCHOSOCIAL STUDY OF CRIMINAL HOMICIDE 2 (1978)).

they would be absent such motives. Given their respective reasons for acting, the resulting purposeful murders that they commit demonstrate less ill will toward their victims than they would absent such strong reasons.[63] Their actions are consistent with them granting some positive value to human life despite a failure in reasoning in allowing that value to be outweighed by their countervailing reasons for engaging in purposeful homicide.

The negligent homicides of Welansky and Chauvin, in contrast, evince a strong failure of reasons-responsiveness. Given the absence of strong countervailing reasons not to inquire, their negligence can be explained only by their failing to ascribe even minimal weight to the value of the lives that their actions endangered. Welansky and Chauvin did not see the lives of their victims as important enough to even consider whether their actions might lead to a risk of their victims' deaths, let alone act to prevent such a risk. And yet, on the PKRN grading regime, Chauvin and Welansky will be graded as least criminally liable on the MPC criminal homicide regime, while the defendants in Browne and William's study will be held to the highest grade of liability in the MPC criminal homicide regime.[64]

This difference in grade translates into substantial consequences for punishment as well. The recommended punishment for negligent homicide in federal jurisdictions, for example, is around one year's imprisonment,[65] with a statutory maximum of eight years' imprisonment.[66] In contrast, purposeful homicide statutes typically involve mandatory minimums of somewhere between several decades' imprisonment to life without parole.[67] While the principle of weak proportionality can withstand some variance in culpability for equally liable defendants,

---

63. For a more dramatic (if more atypical) example, consider again the case of Dr. Kevorkian from Part I. As we saw, Kevorkian intentionally caused the death of his patients, recognizing the prevention of the patient's pain as a strong reason to act. Compare Kevorkian's case to that of the typical reckless or negligent agent, like Chauvin or Welansky, who, for example, fails to take others' potential pain as sufficiently important grounds for inspecting the building to determine whether the fire alarm system needs replacing.

64. For the purposes of this comparative analysis of culpability, I do not account for Chauvin's strict-liability felony homicide conviction. For a more detailed discussion on the exceptional circumstances surrounding Chauvin's felony homicide liability and concerns about strict liability and proportionality generally, see Gideon Yaffe, *The Lucky Legal Accident that Led to Derek Chauvin's Conviction*, HILL (May 1, 2021, 3:00 PM EDT), https://thehill.com/opinion/criminal-justice/551322-the-lucky-legal-accident-that-led-to-derek-chauvins-conviction [https://perma.cc/EWV7-FZC4].

65. *See* U.S. SENT'G GUIDELINES MANUAL § 2A1.4 (U.S. SENT'G COMM'N 2004).

66. 18 U.S.C. § 1112(b) (2018).

67. *See, e.g.*, *id.* § 1111(b).

these cases of dramatically different levels of criminal liability, with higher liability for less culpable defendants, are a clear violation of any plausible principle of proportionality necessary for a justifiable criminal code.

B. Responding to the Argument for Ordinal Convergence

So far, I have argued by counterexample against the equal-culpability thesis on the reasons-responsiveness conception of culpability, which argues that purposeful killings always evince larger (or at least equal) failures of reasons-responsiveness than knowing killings, which evince larger failures of reasons-responsiveness than reckless killings and negligent killings, respectively. I have also suggested that some of these counterexamples involve significant failures of proportionality that challenge the weak proportionality principle to which many proponents of both the MPC and the reasons-responsiveness picture are committed. What I have not yet done, however, is provide a full account of *how* these counterexamples are possible. In particular, I have not addressed head-on the argument for the necessity of ordinal convergence, which I have suggested lies in the background for those who hold that interhierarchical variance is impossible.[68]

Recall the argument for ordinal convergence from Section I.B. If someone harms another person recklessly or knowingly, they acted despite treating the harm (discounted by its probability in cases of recklessness) as a weak reason to *refrain* from acting in their calculus. But if someone causes harm intentionally, they treated the harm as a reason to *pursue* the action. Given that an action that will harm another provides an objectively strong negative reason to refrain, someone who subjectively treats harm as a weak negative reason to refrain deviates from the appropriate response to that reason less so than someone who subjectively treats harm as a positive reason to act. An indifferent agent who treats harm as a reason neither to act nor to refrain from acting lies somewhere in the middle in terms of failures of reasons-responsiveness, since the force they give the reason is less distant from the reason's objective force than the purposeful agent, but further off than the reluctant agent. How, in light of this argument, is it possible for the counterexamples from Section I.B to arise? What do the counterexamples show about where the ordinal-convergence argument went wrong?

The most straightforward response is to deny that the agent who purposefully harms another must treat that harm as a reason to act, rather than as a reason to refrain from acting. An agent who intentionally takes a life need not see that feature of their intentional action as something value-making. Someone

68. *See supra* Section I.B.

may intentionally *A* as a means to *B*, despite seeing *A*, by itself, as disvaluable, or as having features that count as a reason against the entire enterprise. Not every means must be seen as an unalloyed good. Since the wrongful action may be a mere means, the purposeful agent need not see the wrong-making feature of their action as their reason for acting. Indeed, they may see the means as, in and of itself, a regrettable but necessary concomitant of their ultimate goal, just like the knowledgeable agent might see the foreseen consequences of their actions as a regrettable but necessary concomitant of their intended goal.[69] An agent may thus purposefully cause another harm, even though that harm was not, for them, a reason in favor of acting.

This is not to say that the reasons-responsive theorist must deny any distinction between actions performed purposefully and knowingly. If an agent causes a death as a means in some further plan, it is not merely a foreseen consequence. They are committed to taking the life in a variety of ways.[70] Consider the famous trolley case, which compares the person who knowingly causes the death of a bystander by switching the trolley's tracks to save five people who would otherwise have been killed, and the person who intentionally causes the death of the bystander by pushing the bystander onto the tracks as a means to save the five.[71] If it looks like the bystander is escaping in the first scenario, the actor will be relieved. But if it looks like the bystander will escape in the second scenario, then the actor will have to take steps to prevent them from escaping.[72] Still, this normative commitment to ensuring the death of the bystander does not require the purposeful agent to see the fact that the life will be lost, under that description, as a reason counting in favor of the action. They just need to see the consideration that the train will be stopped as a reason counting in favor of the action.[73] Indeed, as philosopher Michael E. Bratman has pointed out, both the agent who acts intentionally (i.e., purposefully) and the agent who acts knowingly might give the lost life the same force in their reasoning as counting against performing the action.[74] After all, both actions will lead to a bystander dying. Given that both

---

69. *See, e.g.*, MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 152-55 (1987) (distinguishing what is chosen on the basis of practical reasoning from what is intended).

70. *Id.* at 155-56.

71. *Cf.* PHILIPPA FOOT, *The Problem of Abortion and the Doctrine of the Double Effect*, *in* VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19, 27-31 (2002) (discussing trolley-like problems and the distinction between direct and oblique intentions).

72. BRATMAN, *supra* note 69, at 156.

73. *Id.* at 152-55.

74. *Id.* at 152. One might reply that the intentional agent must at least see *A*-ing as good, or choice worthy, and thus positive, in a way that the knowing agent need not. This is right, insofar as it goes. But reasons-responsiveness is concerned with the *feature* of *A*-ing by virtue of which

agents factored this consideration into their reasoning about how to act, both agents are equally answerable and so equally normatively committed, despite only one of the two agents intending the result.

The reasons-responsiveness conception of culpability thus creates space for an intentional agent to be less culpable than a reckless agent by focusing less on the question of what state of affairs the agent is committed, through their intentions, to bringing about,[75] and focusing more instead on the question of what features of those states of affairs the agent is committed to evaluating as valuable or disvaluable.[76] Sometimes, these two differing objects of assessment lead to different results about relative culpability.

One might respond that the purposeful agent's normative commitment to pursuing the impermissible result-element as a means will affect the agent's dispositions in ways that necessarily reveal worse reasons-responsiveness to the value of human life than the reckless agent, who merely foresees a possibility of, but does not intend, the impermissible result.[77] After all, a "tracking" disposition to push the fleeing bystander onto the tracks certainly appears to be a larger failure to respond to the value of human life than a disposition not to push the bystander onto the tracks. In this way, the agent's normative commitments to their intended means cannot be severed so simply from an evaluation of their reasons-responsiveness or their quality of will in evaluating the worth of other persons.

But this dispositional response fails to consider equally or more problematic dispositions of the indifferent agent. Consider two agents with the same aim of achieving some goal *G*. One is completely indifferent to human life, and the other finds the value of human life a strong reason to refrain from acting in ways that would result in the loss of such life, though not a strong enough reason to outweigh the value of *G*. Suppose that to achieve *G*, the reluctant agent must push the victim, *V*, onto the tracks, whereas the indifferent agent must drive through *V*, who is on the tracks, foreseeably causing their death. Now suppose a new means opens up. Though more inconvenient, the agents can take some longer path to *G* that neither requires pushing *V* onto the tracks nor driving through *V*. The purposeful agent who values human life will be disposed to take

the agent sees *A*-ing to be good or choice worthy. The intentional agent who *A*s in order to *B* must see *A*-ing as good *qua* being a means to *B*. But they need not see any other feature of *A* (say, that it involves harm to the victim) as good, any more than the agent who foresees but does not intend *A* needs to see that feature of *A* as good. So their seeing *A* as good does not (yet) say anything about their values, until we delve into what they see as good about it, and what features of *A*, so described, they take to be good-making.

75. *See* Gideon Yaffe, *Criminal Attempts*, 124 YALE L.J. 92, 106-14 (2014).

76. *See* WATSON, *supra* note 17, at 131-34.

77. For a discussion of dispositional accounts of reasons-responsiveness, see FISCHER & RAVIZZA, *supra* note 17, at 207, which caches out a form of "guidance control" in terms of the agent's dispositions.

on the inconvenience, and so take alternative means that do not involve causing *V*'s death. In contrast, the indifferent-reckless agent (or negligent agent) who fails to value human life at all would not take themselves to have any reason to adopt the more inconvenient means and so will not be disposed to adopt the more inconvenient means. The reluctant agent, despite pursuing the taking of a human life as a means, has dispositions to respond better in counterfactual cases—dispositions that derive from the fact that the reluctant agent sees the taking of the life as intrinsically disvaluable, and so is ready to take alternative means to avoid the impermissible result in a way that the indifferent agent is not.

C. The Empirical Significance of Interhierarchical Culpability Differences

The possibility for interhierarchical differences in culpability judgments between those operating within a reasons-responsiveness and PKRN conception of subjective culpability also has important consequences for the empirical study of culpability attribution. In a widely cited landmark study, Francis X. Shen, Morris B. Hoffman, Owen D. Jones, and Joshua D. Greene observed that experimental participants' culpability ratings did not map onto the traditional PKRN hierarchy.[78] In a series of experiments, these researchers presented a demographically diverse range of experimental participants with a variety of scenarios, featuring a common fact pattern stem (e.g., "John drops wood planks onto a bike trail, and two bikers crash as a result"), while varying the scenario to make the protagonists' actions purposeful, knowing, reckless, negligent, or "blameless."[79] While the most striking result was that the average relative culpability verdicts assigned to reckless and knowing agents performing the same actions varied repeatedly across scenarios, certain scenarios also showed average culpability verdicts of knowing and reckless agents as more culpable than purposeful agents, and cases of negligent agents judged more culpable than either reckless or knowing agents and nearly identically culpable to purposeful agents.[80]

---

78. *See* Shen et al., *supra* note 16, at 1337-44 (showing some variance across all hierarchies, particularly between knowledge and recklessness). Similar results have been found by Ginther et al., *supra* note 16, at 1330, 1355-60; Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1, 20-22, 25 (2005); and Laurence J. Severance, Jane Goodman & Elizabeth F. Loftus, *Inferring the Criminal Mind: Toward a Bridge Between Legal Doctrine and Psychological Understanding*, 20 J. CRIM. JUST. 107, 111-12, 115 (1992).

79. Shen et al., *supra* note 16, at 1327.

80. *Id.* at 1337-44.

To explain these sorts of startling results, a variety of mechanisms have been formulated. Shen and his colleagues posit that the differences may be due to difficulty on the part of experimental participants in distinguishing between the relevant PKRN mental states.[81] Janice Nadler and Mary-Hunter McDonnell suggest that such culpability attributions may be the result of motivated reasoning, with the desire to punish unlikeable protagonists unconsciously affecting the experimental participant's culpability judgments concerning those protagonist's actions, though they themselves "may not be aware of such influence" and may "regard [this influence] as unjustifiable."[82]

The problem with these studies is that the typical scenario, in varying the agent's PKRN mental states, also varies the agent's reasons for acting as well. While this might be methodologically harmless if a reasons-responsiveness model of subjective culpability led only to intrahierarchical differences in culpability attributions, if a reasons-responsiveness model might also lead to interhierarchical differences in culpability verdicts, then the possibility that participants are operating with such a model, and correctly judging relative culpability according to that model, provides an alternative explanation that the experimental designs have overlooked.

Consider, for example, the variance provided between the purposeful and knowing actor in the illustrative scenario provided by Shen and his colleagues.[83] For the knowing actor, the scenario provided read as follows:

> John is doing carpentry work on his house, which abuts a public mountain bike trail. While carrying wood planks, John drops some onto the trail and doesn't pick them up because he wants to start the carpentry work, even though he is practically certain that in doing so bikers will hit the planks and be injured.[84]

For the purposeful actor, in contrast, participants received an alternative second sentence: "Angry at the mountain bikers for making too much noise biking past his house, one day while carrying a large armload of planks, John desires to injure some bikers and drops some of the planks on to the bike trail."[85]

Notice, first, that in addition to varying the proximate mental state with which John acts (purpose and knowledge), these scenarios also vary the *reason*

---

81. *Id.* at 1352-53. This claim has been tempered in more recent publications, in light of further testing. *See* Ginther et al., *supra* note 16, at 1338.

82. Janice Nadler & Mary-Hunter McDonnell, *Moral Character, Motive, and the Psychology of Blame*, 97 CORNELL L. REV. 255, 270 (2012).

83. Shen et al., *supra* note 16, at 1328.

84. *Id.*

85. *Id.*

for which John acts. In the knowledge case, John acts because of a desire to finish his carpentry; in the purpose case, John acts because of a desire for revenge. Could this difference in reasons result in a case of interhierarchical culpability variance on the reasons-responsiveness picture? Plausibly, yes. In the experimental design, the reason not to act (the harm to the bikers) is identical in both scenarios. Had John given that reason its proper weight, he would have refrained from acting as he did. The question is which John — the knowing or purposeful — weighed that reason less, and so fell shorter from the appropriate reasons-response.

As we have seen, one way for a participant to answer that question is to determine the weight of the countervailing reason. The more subjectively compelling the countervailing reason, the less John needs to have undervalued the reasons that the bikers' well-being gave him to refrain from acting as he did. Thus, participants who rated the knowing agent as more culpable than the purposeful agent may simply have thought that revenge against rowdy bikers is a more subjectively compelling reason for John to act, and so consistent with John's having demonstrated *less* ill will toward the bikers than if he acted to cause them harm merely to finish a personal project when the bikers had done nothing to him. Participants who rated the knowing agent as less culpable, in contrast, may have been responding to the change from purpose to knowledge, or they may have been judging that the carpentry was a more compelling reason for acting than revenge.

In contrast, many of the knowledge and recklessness scenarios provide identical motives for acting. In both the knowledge and the recklessness versions of the illustrative scenario, John acts "because he wants to start the carpentry."[86] The authors are aware that the motives here are underspecified. As they acknowledge, the brevity of the cases "raised a number of questions about how to communicate the protagonist's motivation and intent effectively and efficiently. John's action in each of our scenarios was explained to subjects with a simple, and typically neutral, motivation."[87]

This use of a neutral motivation in a cross-subject study for the purposes of comparing the relative culpability assessments of the knowing or reckless agent is harmless if variances are merely intrahierarchical. Even if different participants read in different strengths to the neutrally presented countervailing reasons within different scenarios, those differences in strength-of-reason should not make a difference in their ultimate culpability judgments across PKRN hierarchies. Given the possibility of interhierarchical differences, however, the dangers

86. *Id.*

87. *Id.* at 1326 n.71.

of providing an underspecified scenario are far greater, particularly in across-subject studies where different participants can read in different details.[88]

Suppose a participant reads in an important construction project to the knowledge scenario, so that the reasons to start the carpentry project are strong. Suppose a second participant reads in a less important construction project to the recklessness scenario, so that the reasons to start the carpentry are weaker. We should expect those two participants to judge the reckless actor to be undervaluing the force of the bikers' well-being as a reason for refraining more than the knowledgeable agent, and so expect them to find the reckless agent more culpable. If we imagine that participants read in the strength of the reason randomly, we should expect at least some variation of when they judge the reckless or knowledgeable agent to be more culpable. Indeed, this is precisely what has been found.[89] Until this possibility is controlled for, there is no need to posit that the difference in responses is due to failures to properly distinguish the mental states of recklessness and knowledge.

III. RECONCILING INTERHIERARCHICAL DIFFERENCES: SOME FIRST STEPS

In this Note, I have explicated two differing pictures of subjective culpability: the PKRN picture, which attributes culpability to the proximate mental states (like an intention) behind an agent's acts, and the reasons-responsiveness picture, which attributes culpability to the more distal mental states (like the reasoning behind the intention to act). I have shown that these two pictures lead to different *intra*hierarchical verdicts about the relative culpability of various reckless or purposeful agents depending on their reasons for recklessly or purposefully acting. I have then considered an attempt to "fit" the MPC's PKRN grading system into a reasons-responsiveness picture of subjective culpability in light of

88. *Id.* at 1324.

89. While I have focused on Shen et al.'s experimental design because their results are most influential, their study design is typical, and a similar problem arises in most similar studies. Consider the experimental designs of Nadler and McDonnell, purporting to show that their subjects' culpability judgments are unconsciously influenced by character likeability through motivated reasoning. Nadler & McDonnell, *supra* note 82. In their experimental design, involving a good-character and bad-character negligent dog owner, they fail to provide the reasons for the activity that caused the dog owner's inattention. *Id.* at 284-88. As they show that experimental participants can identify that the owner neither intended nor foresaw the dog's attack, they assume that the difference in culpability must be unconscious motivated reasoning. *Id.* at 288. However, given the possibility of interhierarchical culpability verdicts, a second possibility is that here, as in the Shen et al. study, participants are simply assuming poorer reasons-responsiveness evinced by the bad character's negligence, and so are correctly (on the reasons-responsiveness picture) assigning more blame.

those intrahierarchical differences by treating the PKRN mental states as proxies for the underlying reasons-responses that are the ultimate source of culpability. I then showed that this attempt to fit the MPC into the reasons-responsiveness picture of subjective culpability is insufficient, because of the existence of *inter*-hierarchical differences in subjective culpability.

What is to be done in light of these interhierarchical differences? The primary purpose of this Note is simply to investigate and clarify the degree to which the reasons-responsiveness and PKRN pictures differ from one another, and to draw out the normative consequences for the MPC. While it is a hope of the Note that such clarity can help promote clear-eyed positive proposals to address these consequences, a full consideration of what such positive proposals would look like is outside the scope of this Note. Still, in this Part, I will briefly survey some possible avenues for positive changes in light of the challenges surveyed here.[90] Despite shortcomings with each solution, I will tentatively suggest that an "absence of ill will" affirmative defense is perhaps the least problematic of the possible solutions canvassed.

One possible response to interhierarchical variance is a complete revamping of the MPC to replace the PKRN mens rea regime at its heart with a new picture of subjective culpability.[91] As we can now see, this response might have more merit than many may have thought in the absence of interhierarchical variance. Such a revamping might be morally required, not only if we are concerned with a "perfect" matching of culpability and liability to which the court may not need to aim, but even to meet minimal requirements of weak proportionality. But such a strategy faces severe downsides. Besides the daunting theoretical challenges of crafting such a code, evidentiary problems with identifying an agent's reasons, political challenges with determining which reasons the law should claim an agent should have taken into account, and inertial challenges against the enormous ramifications such a change would have for every aspect of American criminal law (and well beyond), it would also allow enormous discretion to juries to make decisions of liability, raising serious issues of potential bias.[92]

Another possibility, to avoid jury bias, is to maintain the MPC as is, and accept a less than ideal criminal-law system that allows some failures of proportionality in order to maintain clear, bright-line rules that avoid the bias that

---

90. One possibility is to abandon the attempt to justify criminal law and to hold that the normative acceptability of criminal law is a mistake. Rather, it should be replaced with a more hard-nosed, critical, realist take on criminal law that drops the pretension that the criminal law is, or takes itself to be, normatively justifiable. While there may be something to be said for such a critical take on the criminal law, I will not pursue that option here.

91. *See, e.g.*, ALEXANDER ET AL., *supra* note 31, at 263-88.

92. *See, e.g.*, Samuel R. Sommers & Phoebe C. Ellsworth, *Race in the Courtroom: Perceptions of Guilt and Dispositional Attributions*, 26 PERSONALITY & SOC. PSYCH. BULL. 1367 (2000).

creeps in when juries or other actors are given more discretion. This is, of course, the avenue that the standard "mens rea as culpability proxy" picture endorses. What this Note has shown, however, is that a clear-eyed picture of the relationship between the reasons-responsiveness and PKRN pictures of subjective culpability reveals that the extent of disproportionate liability is far higher than most theorists who choose this path have acknowledged. The trade-offs between the threats to proportionality posed by increased discretion on the one hand, and the rigidity of the PKRN mens rea system on the other, are substantially larger than has been assumed.

In light of such trade-offs, a third possibility is to seek a less dramatic amendment to the MPC that preserves the PKRN mens rea regime, while going at least some way towards mitigating its worst tendencies toward interhierarchical variance. Luckily, at least some less dramatic possibilities do suggest themselves, by building on the methods the MPC already employs to deal with intrarhierarchical differences, as described in Section I.D. As we have seen, one way that the MPC deals with intrahierarchical differences within a grade is by bringing culpability to bear at the sentencing stage, rather than in the initial assignment of criminal liability. Between two defendants both convicted of, for example, reckless homicide, differences in motive, reasons, or quality of will might be taken into account by the judge in determining sentencing.

One possibility for addressing the existence of interhierarchical differences would be to increase this judicial discretion. The MPC already treats most ungraded crimes as requiring a default recklessness mens rea standard, with purposeful and knowing agents all equally criminally liable, and the judge factoring in differences in culpability at sentencing.[93] If intentional agents are sometimes less liable (because they act with less ill will) than reckless agents, this recklessness-plus standard might let judges correct for those differences. Thus, one solution would be simply to abolish the PKRN grading of crimes like homicide and allow judges sentencing discretion to ensure that less culpable intentional killings are not treated as more serious than more culpable reckless killings. Since reckless agents might be routinely more culpable than knowledgeable agents or purposeful agents for some crimes, this recklessness-plus schema is preferable to the knowledge-plus scheme currently treated as the interpretive default in federal criminal law.[94] Since, as we have seen, even negligent actors can be more

93. *See* MODEL PENAL CODE § 2.02(3) (AM. L. INST. 1962).

94. *See* Morissette v. United States, 342 U.S. 246, 256-58 (1952); Liparota v. United States, 471 U.S. 419, 425 (1985); United States v. X-Citement Video, Inc., 513 U.S. 64, 70-73 (1994); Staples v. United States, 511 U.S. 600, 605-06 (1994); Rehaif v. United States, 139 S. Ct. 2191, 2196 (2019) ("[O]ur reading [of a default knowledge requirement] is consistent with a basic principle that underlies criminal law, namely, the importance of showing what Blackstone

culpable than purposeful actors, it may be that a purely undifferentiated mens rea scheme, where many crimes require mere negligence and judges respond to interhierarchical differences in subjective culpability at the sentencing phase, would best fit the reasons-responsiveness picture.

The downside, of course, is that just as with the strategy of putting an evaluation of an agent's reasons in the jury's hand by making it part of the mens rea elements in the prima facie case, pushing it into the sentencing stage creates an analogous problem of judicial bias. While there may be some empirical evidence that judges are marginally better suited to such tasks than juries, there is substantially more empirical evidence that allowing judges more discretion to shape punishment based on the quality of the offender's perceived motive might let in implicit biases.[95]

A second way the MPC deals with intrahierarchical differences is with mens rea "bump-ups," such as those for committing an act "recklessly under circumstances manifesting extreme indifference to the value of human life," which seem to build in the reasons of the agent, or purpose with "extreme mental or emotional disturbance."[96] So far, these differences typically serve simply to treat marginal cases of more culpable reckless actors as equivalent to knowing and purposeful actors (or the least culpable purposeful and knowing actors as equivalent to reckless actors). But we could imagine a "chutes-and-ladders" version of the MPC where such bump-ups and bump-downs become more frequent, and where they can move a defendant further up or down in criminal grade. For example, purposeful killing may, with sufficient lack of ill will, get bumped down

called 'a vicious will.' As this Court has explained, the understanding that an injury is criminal only if inflicted knowingly 'is as universal and persistent in mature systems of law as belief in freedom of the human will . . . .' Scienter requirements advance this basic principle of criminal law by helping to 'separate those who understand the wrongful nature of their act from those who do not.'" (citations omitted)). Notice that the Court in *Rehaif* explicitly ties its knowledge requirement (that is, scienter) to the hierarchical view that knowledge is an expression of a more "vicious will" than mere recklessness or negligence. If the arguments in Part II are correct, then this view is a mistake.

95. *See, e.g.*, Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007); Jeffrey J. Rachlinski, Sheri Lynn Johnson, Andrew J. Wistrich & Chris Guthrie, *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195 (2009); *cf.* Brian Nosek & Rachel G. Riskind, *Policy Implications of Social Cognition*, 6 SOC. ISSUES & POL'Y REV. 113, 113 (2012) (studying "the present state of evidence for implicit social cognition and its implications for social policy"); Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. PERSONALITY & SOC. PSYCH. 597 (2006) (discussing the larger effects of implicit bias on jurors for similar discretionary tasks involving culpability attribution).

96. MODEL PENAL CODE §§ 210.2(1)(b), 210.3(1)(b) (AM. L. INST. 1962).

to be categorized with negligent homicide, and a negligent killing that demonstrates sufficiently extreme indifference to human life can get bumped up to be categorized, along with knowing killings, as murder. As with hate crime legislation, certain specific patterns of reasons-response that societies judge to be particularly culpable could also be enshrined in the criminal code through legislation for enhanced penalties.[97]

Again, however, this method has its limitations. In particular, a piecemeal approach may find itself hemmed in by the twin problems of rigidity and biased discretion. Insofar as the added mens rea categories are narrowly constrained, they risk missing important cases of interhierarchical variance and so allowing more substantial failures of proportionality. And insofar as the added mens rea categories are broader in scope, they risk allowing jury bias to creep in with the increased discretion juries must apply in making more sophisticated mens rea determinations.

A third and final way that the MPC deals with intrahierarchical differences in culpability is with affirmative defenses. This method is, I tentatively suggest, the most promising. While affirmative defenses will not allow for increasing the liability of the highly culpable negligent or reckless agent, they might serve to address the more problematic case of unusually nonculpable purposeful actors. One possible affirmative excusing defense is "non-ill will," similar to the affirmative defense of duress,[98] which does not justify the action, but can suffice to bump down the offense and limit the criminal liability involved. This method has the added benefit of placing the burden of proof for an "excusing motive" on the defendant, rather than on the prosecutor, which may ease the concerns of those worried that a motive-based criminal-law system would be unmanageable, given the likely difficulty in proving motive mens rea elements. While it allows for the introduction of some bias on the part of a jury, as biased juries might still hold more culpable agents less liable under such a defense for pernicious reasons, it would prevent the more serious problem of biased juries or judges holding less culpable defendants more criminally liable.

---

97. *See, e.g.*, CAL. PENAL CODE § 422.55(a) (West 2021) ("'Hate crime' means a criminal act committed, in whole or in part, because of one or more of the following actual or perceived characteristics of the victim: (1) Disability. (2) Gender. (3) Nationality. (4) Race or ethnicity. (5) Religion. (6) Sexual orientation. (7) Association with a person or group with one or more of these actual or perceived characteristics."). For further discussion, see FREDERICK M. LAWRENCE, PUNISHING HATE: BIAS CRIMES UNDER AMERICAN LAW (1999).

98. MODEL PENAL CODE § 2.09(1) (AM. L. INST. 1962).

CONCLUSION

In this Note, I have explored how we might fit the MPC into a reasons-responsiveness picture of subjective culpability for which it was not initially designed. I have argued that three commitments common to many criminal-law theorists are jointly incompatible: (1) a commitment to a weak proportionality principle according to which a minimally acceptable criminal system must not hold substantially less culpable agents substantially more criminally liable than substantially more culpable agents for the same criminal act; (2) a commitment to the view that the PKRN mens rea regime laid out in the MPC, according to which criminal liability is a function of the PKRN mens rea elements, is at least minimally acceptable; and (3) a commitment to a reasons-responsiveness conception of subjective culpability according to which the locus of an actor's culpability lies not in their purpose, knowledge, recklessness, or negligence, but rather in the responsiveness of the agent's reasoning capacities, which an agent's actions, given the agent's purpose, knowledge, recklessness, or negligence, evince. The fact that the mental states involved in the reasons-responsiveness conception of culpability differ from the mental states of the PKRN hierarchy opens the possibility of misalignment between an agent's culpability and their criminal liability, creating pressure on the commitment to proportionality between the two.

I have further argued that such misalignment will predictably occur for certain important criminal offenses, such as criminal homicide, between cases of reluctant purposeful agents and indifferent reckless or negligent actors. I have shown that reluctant purposeful agents demonstrate less ill will toward their victims, and so are less culpable on a reasons-responsiveness picture of culpability, than are indifferent reckless or negligent agents who commit the same offense. The resulting disproportionate assignments of criminal liability by the MPC should be deeply troubling to those criminal-law theorists who share the three commitments laid out above.

Finally, I have suggested that this potentially dramatic misalignment between culpability and liability across levels of the MPC mens rea hierarchy may help explain certain popular responses to the criminal-justice system that have puzzled criminal-law theorists. The existence of interhierarchical variance can provide an alternative and compelling explanation for recent empirical results concerning experimental subjects' sorting of traditional PKRN mens rea that run counter to the MPC model's expectations. Researchers have overlooked the possibility that experimental findings about study participants' culpability reports that ran counter to the PKRN grading scheme, which researchers thought rationally unjustifiable, and thus attributed to cognitive biases or lack of conceptual understanding of jury instructions, might actually be better explained by the

hypothesis that those participants are operating with a coherent reasons-responsiveness model of culpability, leading them to interhierarchical culpability verdicts.

If this Note is correct, it suggests a new way to understand the popular disconnect between lay outrage against cases of police negligence such as the killing of George Floyd and the often-minimal legal consequences for those killings. The reason for the disparity is not simply a hesitancy on the part of prosecutors to enforce existing criminal statutes against police officers or the existence of specialized shields from liability for police use of force. Built into our general system of liability is the assumption that the actions of an agent who causes some harm intentionally are more culpable than the actions of an agent who causes the same harm unintentionally, where the harm—or potential risk of harm—is merely foreseen. If this Note is correct, then that assumption should be far more controversial than has typically been assumed. The cases of potential misalignment between criminal liability and culpability for reluctant purposeful agents and indifferent negligent agents highlighted in this Note suggest the need for more research on—and perhaps a broader reevaluation of—the prominent role of intention in criminal law more generally.