

8-1-2013

Intentional Killing Without Intending to Kill: Knobe's Theory as a Rational Limit on Felony Murder

Joseph C. Mauro

Follow this and additional works at: <https://digitalcommons.law.lsu.edu/lalrev>



Part of the Law Commons

Repository Citation

Joseph C. Mauro, *Intentional Killing Without Intending to Kill: Knobe's Theory as a Rational Limit on Felony Murder*, 73 La. L. Rev. (2013)

Available at: <https://digitalcommons.law.lsu.edu/lalrev/vol73/iss4/6>

This Article is brought to you for free and open access by the Law Reviews and Journals at LSU Law Digital Commons. It has been accepted for inclusion in Louisiana Law Review by an authorized editor of LSU Law Digital Commons. For more information, please contact kreed25@lsu.edu.

Intentional Killing Without Intending to Kill: Knobe's Theory as a Rational Limit on Felony Murder

*Joseph C. Mauro**

ABSTRACT

Felony murder authorizes maximum criminal punishment, the kind typically reserved for the most ruthless and calculating killers, for defendants who did not even intend to kill. Many retributivist scholars therefore criticize felony murder for abandoning the traditional notion that intent determines culpability and the appropriate degree of punishment. Despite widespread agreement with this criticism, however, felony murder persists in most jurisdictions.

Joshua Knobe's empirical research turns this problem on its head by suggesting that intent is not just a mental state. Numerous experiments have shown, in fact, that people are more likely to call an action intentional, regardless of what the actor is thinking, when it involves morally bad conduct and outcomes. Knobe thus argues that intent refers not only to a mental state but also to the morality (good or bad) of conduct and the outcomes it causes.

Under Knobe's theory, some instances of felony murder are actually intentional, even if the defendant did not subjectively intend to kill or recklessly endanger. That is, when the defendant's conduct (a dangerous felony), mental state (willingness to risk death) and the outcome (death) are sufficiently bad, the action, as a whole, is intentional. Restricting felony murder to such "intentional" killings, I propose, would not only re-couple punishment to intent, satisfying retributive theory, but also eliminate the most troublesome applications of the rule.

Copyright 2013, by JOSEPH C. MAURO.

* Associate, Arent Fox LLP. I owe many thanks to Thomas A. Green, Lawrence Ngo, Walter Sinnott-Armstrong, Samuel R. Gross, and Phoebe C. Ellsworth for their invaluable guidance and comments. I would also like to thank the *Louisiana Law Review* staff for its hard work editing this article. Finally and as always, I am exceedingly grateful to my family and friends for their constant love and support.

TABLE OF CONTENTS

	Abstract	1011
I.	Introduction	1012
II.	The Retributivist Criticism: Felony Murder Ignores Gradations of Culpability	1015
III.	Argument: Felony Murder Can Be Limited to Intentional Killings and Thus Justified Under Retributive Theory	1018
	A. Knobe's Theory: Intent Describes Morality	1019
	B. Retributive Theory: Intent Justifies Murder-Level Punishment.....	1025
	C. Restricting Felony Murder to Intentional Killings.....	1029
IV.	Counterarguments	1034
	A. Felony Murder Is Especially Corrupt and Unfair	1034
	B. Knobe's Theory Does Not Apply to Unforeseen Outcomes	1036
	C. The Knobe Effect Represents a Mistaken Perception	1039
	1. Hindsight Bias	1040
	2. Motivational Bias	1042
	3. Outcome Bias	1044
V.	Conclusion.....	1047

I. INTRODUCTION

Consider this scenario. The CEO of a major corporation has developed a new business plan. Her only goal is to make money, and the plan is going to make lots of it. Before the plan is implemented, one of her advisors comes to her looking distressed. "Your plan is going to destroy 200,000 acres of rainforest," says the advisor.

“I’ve never cared about the environment,” says the CEO. “I only care about money. Now go ahead with the plan.”

The CEO of a second corporation has developed a new plan as well. Like the first CEO, her only goal is to make money, and her plan is going to make lots of it. Before the plan is implemented, one of her advisors comes to her looking excited. “Not only is your plan going to make lots of money,” he says, “it’s also going to save 200,000 acres of rainforest that otherwise would have been destroyed.”

“I’ve never cared about the environment,” says the CEO. “I only care about money. Now go ahead with the plan.”

Did the CEOs intentionally affect the environment? In a sense, neither of them did. They only cared about money and had no feelings about the environment. On the other hand, both CEOs intentionally affected the environment. They knew that their actions would affect the environment and decided to act. Yet whether they intentionally affected the environment or not, the answer must be the same for each CEO, right?

Surprisingly, most people give a different answer for each CEO. When asked about these scenarios, 87% of people said the first CEO intentionally harmed the environment, while only 20% said the second CEO intentionally helped the environment.¹ Numerous studies have obtained similar results, and one trend among the studies is that when people perceive an action to be morally bad, they are more likely to call it intentional. This phenomenon, broadly speaking, is called the Knobe Effect.²

According to some experts, this research suggests that *intent* refers not only to a person’s mental state but also to the morality of her conduct and the outcomes it causes.³ Under this theory, when conduct, outcomes, and the person’s mental state are sufficiently bad, the action is intentional.⁴ Some experts disagree, maintaining

1. Joshua Knobe & Arudra Burra, *The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study*, 6 J. COGNITION & CULTURE 113, 117–18 (2006).

2. Some researchers define the Knobe Effect more narrowly, applying it only to Joshua Knobe’s 2003 study. Research involving the phenomenon has expanded since 2003, however, and this Article uses the term *Knobe Effect* to refer to the broader phenomenon that the morality of an action affects people’s perceptions of whether it was intentional. See *infra* Part II.A.

3. Knobe & Burra, *supra* note 1, at 123–25. The same is true of words such as *intention*, *intentionally*, and *intend*.

4. *Id.* See also Thomas Nadelhoffer, *Skill, Luck, Control, and Intentional Action*, 18 PHIL. PSYCHOL. 341, 351 (2005) (“[B]y surveying the folk, I hoped to find out which conceptual analyses of intentional action actually settle with their intuitions, so that philosophers will no longer be able to align their analyses with common sense unless their views empirically merit such support. Of course, that

that intent is nothing but a mental state and that the Knobe Effect simply represents a mistaken perception of mental states.⁵ Nevertheless, Joshua Knobe, the author of the study involving the CEOs, argues that the Knobe Effect is relevant to the meaning of intent.⁶ An action can be intentional, he says, by virtue of morally bad conduct and outcomes, even if the person did not subjectively want the outcomes to occur. Knobe's is a more objectively observable conception of intent.⁷

Knobe's theory casts new light on the most common retributivist criticism of felony murder. Felony murder authorizes murder-level punishment when the defendant killed in the course of committing a dangerous felony, even if the defendant did not have the mental state traditionally required for murder.⁸ Whereas murder typically requires proof of a certain mental state—for example, “premeditation and deliberation,” “intent to kill,” or “recklessness”—felony murder requires none of these things. In cases of felony murder, most jurisdictions simply presume that the defendant had the mental state required for murder because a dangerous felony caused a person to die.¹⁰

Retributivist critics therefore complain that felony murder does not tailor punishment to the defendant's culpability.¹¹ Intent determines culpability, they say, and felony murder is thus

would not mean that analyses privileged by the endorsement of the folk are *true*, only that the burden of proof would be placed squarely on the shoulders” of the opponents of such analyses.)

5. Nadelhoffer, *supra* note 4, at 341, 344. See also Joshua Knobe, *Person as Scientist, Person as Moralizer*, 33 BEHAV. & BRAIN SCI. 315 (2010).

6. Knobe & Burra, *supra* note 1, at 123–25.

7. Knobe's theory of intentional action still requires the person to have a morally bad mental state regarding the outcome but does not require the person to subjectively intend to bring it about or even risk it.

8. 40 C.J.S. *Homicide* § 51 (2006).

9. *Id.* §§ 40, 42.

10. See Nelson E. Roth & Scott E. Sundby, *The Felony-Murder Rule: A Doctrine at Constitutional Crossroads*, 70 CORNELL L. REV. 446, 449, 455–58 (1985). The felony's dangerousness is important because most states restrict felony murder to underlying felonies that are clearly or inherently dangerous, with the idea that, in such cases, the death was foreseeable. See Leonard Birdsong, *Felony Murder: A Historical Perspective by Which to Understand Today's Modern Felony Murder Rule Statutes*, 32 T. MARSHALL L. REV. 1, 21 & n.140 (2006). This limitation is important for Knobe as well because if a killing is purely accidental, people will not make the moral judgment that causes them to label it intentional. See *infra* Parts II.C, III.A.

11. Felony murder is attacked under other theories of punishment as well, but those arguments are beyond the scope of this Article, which is confined to the retributivist framework.

illegitimate because it does not require proof of intent.¹² Without proof of intent, the state cannot judge whether the defendant is culpable enough for murder-level punishment.¹³

Knobe's theory addresses this problem by suggesting that lawmakers can actually amend the felony murder rule to require proof of intent.¹⁴ Even if the defendant did not mean to kill, the killing was intentional under Knobe's theory if the defendant's mental state and conduct were sufficiently immoral, taking for granted that the outcome of death is maximally immoral.¹⁵ Moreover, if felony murder were restricted to such cases, then it would be better tailored to the most culpable defendants and would no longer fall subject to the retributivist charge that it ignores intent. Part I articulates this retributivist criticism, Part II describes how Knobe's theory addresses it, and Part III deals with three counterarguments.

II. THE RETRIBUTIVIST CRITICISM: FELONY MURDER IGNORES GRADATIONS OF CULPABILITY

Retributivist scholars criticize felony murder for departing from the traditional framework in which intent determines culpability and culpability determines punishment.¹⁶ Felony murder, they say,

12. David Crump, *Reconsidering the Felony Murder Rule in Light of Modern Criticisms: Doesn't the Conclusion Depend upon the Particular Rule at Issue?*, 32 HARV. J.L. & PUB. POL'Y 1155, 1159–60 (2009); James J. Tomkovicz, *The Endurance of the Felony-Murder Rule: A Study of the Forces That Shape Our Criminal Law*, 51 WASH. & LEE L. REV. 1429, 1441 (1994). Although states use various sentencing schemes to determine which punishments are appropriate, a murder conviction almost always authorizes greater punishment than a conviction for negligent homicide and often authorizes the most severe punishment in the jurisdiction. See 40 C.J.S. *Homicide* §§ 73, 75.

13. Crump, *supra* note 12, at 1159–60. But see *infra* Part II.B (discussing whether subjective intent is the only component of culpability).

14. I do not mean to suggest that the concept of legal intent is uniform. To the contrary, legal doctrines vary with respect to how they define and use the concept of intent. See, e.g., *McCleskey v. Kemp*, 481 U.S. 279, 293–95 & n.12 (1987) (distinguishing the intent elements of Title VII cases, capital cases, and voting rights cases). My argument simply focuses on the *linguistic* meaning of intent, i.e., what people actually mean when they talk about it. I argue that in the homicide context, the linguistic meaning of intent is crucial because a murder conviction is not justified (under the retributivist theory) unless the killing was *truly* intentional, not just legally “intentional.”

15. See *infra* Part II.C for a discussion of which types of conduct and mental states can make an undesired killing intentional.

16. See PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY & BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* 170–71 (1995). See also generally Norman J. Finkel & Stefanie F. Smith, *Principals and Accessories in Capital Felony-Murder: The Proportionality Principle Reigns Supreme*, 27 LAW & SOC'Y REV. 129 (1993).

ignores gradations of culpability by not requiring proof of intent.¹⁷ In other words, murder convictions should require proof that the defendant had a highly culpable intent regarding the death—that she intended it or exhibited a depraved recklessness toward it¹⁸—and felony murder ignores this requirement. Many retributivist scholars therefore consider the rule unjustified.¹⁹

Indeed, felony murder is the only murder rule that does not require proof of any particular mental state.²⁰ Traditionally, a homicide is not murder unless the state proves that it was done with “premeditation and deliberation,” “recklessness,” “malice aforethought,” “indifference to human life,” or any number of other highly culpable mental states.²¹ In cases of felony murder, on the other hand,

the only intent required for a killer to be convicted . . . is generally the intent to commit or to participate in the underlying felony, and felony murder generally does not require as elements intent to kill, malice, premeditation or deliberation, willfulness, or even that [the] defendant [knew]

17. Crump, *supra* note 12, at 1159–60; Tomkovicz, *supra* note 12, at 1441.

18. Tomkovicz, *supra* note 12, at 1439–40.

19. That is, unless one accepts the “constructive malice theory” or the “evil mind theory.” See Roth & Sundby, *supra* note 10, at 455–58. The “evil mind” theory presumes that felony murder defendants satisfy the intent element of murder, whatever it may be in the jurisdiction, because the felonious action that caused the death proves that the defendant had an “evil mind,” which is sufficient evidence of culpable intent. *Id.* The “constructive malice” theory, on the other hand, does not *presume* that the defendant had a culpable intent but holds instead that the underlying felony, regardless of what was going on in the defendant’s mind, makes her so culpable that the state is justified in *imputing* the mens rea for murder to her. The “constructive malice” rationale is thus intellectually honest about the fact that the defendant may not have had a subjective intent to kill or recklessly endanger and therefore employs “a ‘mens rea-imposing mechanism’” that treats the defendant *as though* she had the subjective intent traditionally required for murder. *Id.* at 455–56 (citation omitted).

20. Compare 40 C.J.S. *Homicide* § 51 (2006), with 40 C.J.S. *Homicide* §§ 34, 36, 38, 39. Even so, 35 states restrict felony murder to “inherently” dangerous felonies that show that the defendant at least acted with some amount of carelessness or disdain for human life. See Birdsong, *supra* note 10, at 21 & n.140; Crump, *supra* note 12, at 1181 (explaining that the “primitive form of felony murder [under which, for example, a pickpocket can be convicted of murder when his mark dies in a freak accident] is a caricature, inasmuch as modern jurisdictions confine the felony murder rule [to inherently dangerous felonies or conduct]” (citation omitted)). See also Guyora Binder, *The Origins of American Felony Murder Rules*, 57 STAN. L. REV. 59, 202 (2004) (observing that American felony murder laws, from their inception, have been confined to felonies involving “the deliberate infliction of violence”).

21. 40 C.J.S. *Homicide* §§ 40, 42.

that his or her acts [would] cause death or great bodily harm to the victim.²²

Additionally, the state does not need to prove the underlying felony. The underlying felony is simply a factual element of the crime: “[A] felony-murder conviction may stand even if the underlying felony which serves as its predicate is not submitted to the jury, or is dismissed, or there is an acquittal of the underlying felony, or where a conviction on the underlying felony is vacated.”²³

A number of courts and scholars have articulated the above criticism of felony murder.²⁴ The Model Penal Code, for example, appears to have abolished felony murder because it lacks an intent element.²⁵ James Tomkovicz similarly argues that felony murder is

22. *Id.* § 51.

23. 40 C.J.S. *Homicide* § 51. A few states, however, do not allow convictions of first-degree felony murder without proof of malice or premeditation and deliberation. *Id.* n.29–32.

24. See Crump, *supra* note 12, at 1158 n.10 (listing several articles advancing this criticism, including Robert M. Elliot, *The Merger Doctrine as a Limitation on the Felony-Murder Rule: A Balance of Criminal Law Principles*, 13 WAKE FOREST L. REV. 369, 371 (1977) (“[T]he rule does violence to the philosophy which dictates that criminal liability should be commensurate with moral culpability.”); George P. Fletcher, *Reflections on Felony-Murder*, 12 SW. U. L. REV. 413, 428 (1981) (“When the felony-murder rule converts an accidental death into first-degree murder, then punishment is rendered disproportionate to the wrong for which the offender is personally responsible.”); Jeanne H. Seibold, *The Felony-Murder Rule: In Search of a Viable Doctrine*, 23 CATH. LAW. 133, 160 (1978) (“The concept of basing the degree of punishment on the seriousness of the result of the criminal act seems grossly misplaced in a legal system which recognizes the degree of mental culpability as the appropriate standard for fixing criminal liability.”)).

25. *Id.* at 1159–60, 1164 (citing MODEL PENAL CODE § 210.2 cmt. (1982)). On the other hand, the Model Penal Code (MPC) appears to have crafted its definition of homicide to make room for felony murder by allowing murder convictions when the defendant exhibited “extreme indifference” to human life. *Id.* at 1164 (“[T]he MPC attempts to cover the felony murder situation by an odd combination: a confusing concept of recklessness coupled with a presumption. First, the MPC defines murder to include homicides committed ‘recklessly under circumstances manifesting extreme indifference to the value of human life.’ This sentence contains several vague concepts that are likely to produce inconsistency and arbitrariness in verdicts. For example, two juries can differ significantly enough over the kinds of indifference that are ‘extreme’ that they produce seriously inconsistent verdicts. Furthermore, some jurors may be motivated to find ‘extreme’ indifference because of improper factors such as the defendant’s lifestyle, personality, or ethnicity. Next, the MPC provides that the requisite recklessness and indifference are ‘presumed’ if the actor was engaged in any of several named felonies. A criminal presumption, of course, requires the judge to tell the jury that it can follow or disregard the presumption as it chooses. A well-crafted felony murder law would provide greater clarity and thus confine discretion better than the MPC’s backdoor method of ostensibly ‘abolishing’ the

unjustified because it does not require proof of the defendant's mental state.²⁶ Retributive criminal theory, he says, depends upon "the idea that punishment should be commensurate with *mental* fault."²⁷ The Michigan Supreme Court articulated the same criticism in abolishing felony murder in that state:

It is not sound principle to convert an accidental, negligent, or reckless homicide into a murder simply because, without more, the killing was in furtherance of a criminal objective of some defined class. [Instead, murder requires a showing of] recklessness . . . or a practical certainty or intent, with respect to causing death, [which is] an independent determination . . . on the facts of each case.²⁸

The court is likely correct that, "without more," a killing during the commission of a felony is insufficient to prove murder-level culpability. Traditional retributive theory requires, at a minimum, some proof of intent. Accordingly, to bring felony murder in line with retributive theory, it would need to be limited to cases in which the killing was actually intentional.

III. ARGUMENT: FELONY MURDER CAN BE LIMITED TO INTENTIONAL KILLINGS AND THUS JUSTIFIED UNDER RETRIBUTIVE THEORY

According to Knobe, the Knobe Effect shows that intent refers not only to a mental state but also to the morality of conduct and outcomes.²⁹ If this is true, then some cases of felony murder involve intentional killings because the immoral conduct (a dangerous felony), immoral outcome (a death), and a sufficiently immoral mental state regarding the death (a willingness to risk it), combine to make the action, as a whole, intentional.³⁰ In fact, the indicators of intent under Knobe's theory (conduct, outcomes, and mental state) are the same factors that some retributivist scholars would use to tailor felony murder to the most culpable defendants.³¹ A properly

rule while actually preserving it in an altered form." (citations omitted)). *See also infra* Part III.A (discussing the similarity between the Michigan Supreme Court's "abrogation" of felony murder in *People v. Aaron* and a restriction to killings that are intentional under Knobe's theory).

26. Tomkovicz, *supra* note 12, at 1441.

27. *Id.* (emphasis added).

28. *People v. Aaron*, 299 N.W.2d 304, 314 (Mich. 1980) (citation omitted).

29. Knobe & Burra, *supra* note 1, at 123–25.

30. *See infra* Part III.C (discussing which mental states are required to call a killing intentional under Knobe's theory).

31. *See infra* Part II.B.

limited version of felony murder, therefore, far from undermining intent-based culpability, would actually support it.

A. Knobe's Theory: Intent Describes Morality

Following Knobe's groundbreaking study, subsequent research confirmed and expanded his findings.³² For example, studies found that the Knobe Effect occurs when intent is described in terms such as *intention*, *deciding*, *desire*, *in favor of*, *in order to*, and *advocating*.³³ It also occurs at what might be called the edges of linguistic intent: In the scenario involving the CEOs, for instance, people are willing to say that the first CEO "increased profits by harming the environment" but not that the second CEO "increased profits by helping the environment."³⁴ Another study showed that the Knobe Effect occurs even when the actor did not know that the bad outcome would result but nonetheless wanted it to happen.³⁵ Thus, for the purposes of this Article, the *Knobe Effect* refers to the entire body of research showing that the morality of an action affects the extent to which it is, or is perceived to be, intentional.

According to Knobe, this research shows that intent is more than a mental state. He argues that *subjectively intending* something is distinct from the broader concept of *acting intentionally*,³⁶ which describes the morality of conduct and outcomes in addition to mental states.³⁷ In other words, immoral conduct that causes a bad outcome can make an action intentional even in the absence of

32. Recall that Knobe's study about the CEOs revealed that people are more likely to call morally bad actions intentional. Knobe & Burra, *supra* note 1, at 118–19. When the CEO harmed the environment, 87% of people said she did so intentionally, but when the CEO helped the environment, only 20% of people said she did so intentionally. *Id.*

33. Knobe, *supra* note 5, at 318; JOSHUA KNOBE & SHAUN NICHOLS, EXPERIMENTAL PHILOSOPHY 137 (2008). Additionally, one study showed that "people are willing to say: 'The chairman harmed the environment in order to increase profits. But not: the chairman helped the environment in order to increase profits.'" Knobe, *supra* note 5, at 319.

34. *Id.*

35. Nadelhoffer, *supra* note 4, at 345–46. *See also infra* Part III.B (discussing the possibility that the Knobe Effect does not occur unless the person *knew* her actions would cause the relevant outcome).

36. Knobe & Burra, *supra* note 1, at 124.

37. *But see* Knobe, *supra* note 5, at 349 (referencing Thomas Scanlon's argument that the Knobe Effect actually depends on people's perceptions of mental states). *See also infra* Part III.C (discussing whether the Knobe Effect alters the meaning of intent (Knobe's theory) or simply represents a mistaken perception of mental states).

subjective intent.³⁸ If Knobe is correct, then felony murders are often intentional because, when a dangerous felony causes a death, the moral badness of the conduct and outcome make the entire action intentional even if the person did not subjectively intend to kill.³⁹

Lawrence Solan describes the Knobe Effect's mechanism, illuminating the connection between Knobe's theory and felony murder. He hypothesizes a baseline of subjective intentions that people expect of one another.⁴⁰ In general, people assume that everyone intends good outcomes.⁴¹ If a person fails to intend good outcomes—for example, when the CEO does not care about affecting the environment—then that person falls below the baseline of expected intentions.

Solan thus argues that only morally bad actions need to be described as intentional because people assume, based on the baseline, that good actions were subjectively intended.⁴² He refers to Paul Grice's "Maxim of Quantity," which holds that dialogue should only include relevant statements, and concludes that people need not mention the intent behind an action unless it contradicts the general presumption that everyone intends good outcomes.⁴³ For purposes of this article, Solan's theory of the Knobe Effect's mechanics will be called the *baseline theory*.⁴⁴

38. According to Knobe, there is no reason to assume that "a theory of intentional action needs to be related to a conception of future intention." Knobe, *supra* note 1, at 115.

39. *Id.* Again, this assumes that the defendant had some culpable mental state, short of intent to kill or recklessly endanger, regarding the death. *See infra* Part II.C. *But see* Lawrence M. Solan, *Blame, Praise, and the Structure of Legal Rules*, 75 BROOK. L. REV. 517, 526–27, 529 (2009) (discussing Steve Guglielmo and Bertram Malle's explanation of the Knobe Effect and referring to a study by Steven Sverdlik that found that "when given the opportunity, people are perfectly willing to find conduct both blameworthy and unintentional at the same time" (citation omitted)); Knobe, *supra* note 5, at 339 (referencing Frank Hindriks's theory that what motivates the Knobe Effect is the actor's failure to take adequate precautions to avoid a foreseen bad outcome: "In criminal law, foresight betrays a guilty mind as much as intent does: both reveal that the agent is not properly motivated to avoid an illegal state of affairs. This commonality warrants our judgment that the state is brought about intentionally, even when unintended. In contrast to Knobe, I thus retain the idea that acting intentionally is acting with a certain frame of mind.").

40. There may be a normative element to these expectations, but as described in Solan's piece, they are primarily predictive. *See* Solan, *supra* note 39, at 522–28.

41. *Id.*

42. *Id.* at 522.

43. *Id.*

44. Joseph Ulatowski and Justus Johnson contend that while it might make sense to talk about a baseline of expected intentions, the theory is not valid unless

Solan proves his baseline theory with a thought experiment involving Mario and a printer.⁴⁵ When someone says, “Mario fixed the printer,” there is no need to say he did so intentionally because people presume that he intended to cause the good outcome. In contrast, when someone says, “Mario broke the printer,” people assume that he did not intend to do so because the outcome was bad. Thus, neither of these scenarios gives people a reason to call Mario’s actions intentional because there is no reason to think that he fell below the baseline of expected intentions.

People abandon their expectations, however, when they learn that sub-baseline intentions caused morally bad outcomes—for example, when the CEO did not care about the environment and then harmed it. Unlike Mario, the CEO fell below the perceived baseline of expected intentions because she did not care about harming the environment. When the environment was harmed, people therefore had a reason to say something about her intent. People noted her sub-baseline intentions, and expressed disapproval of them,⁴⁶ by saying that she caused the harm intentionally.⁴⁷ Yet, importantly, sub-baseline

the baseline is more clearly defined. Knobe, *supra* note 5, at 352–53 (“[W]e find it hard to ascribe a coherent view to some experimental subjects if the default position is not clearly defined.”). Additionally, some research suggests that morality and culpability are unrelated to the perception of intent. *See, e.g.*, Knobe, *supra* note 5, at 349 (wherein Thomas M. Scanlon argues that Knobe’s description of the Knobe Effect equivocates between being “*belief-intentional*” (believing the act is being performed) and “*aim-intentional*” (believing the outcome will happen), which explains the Knobe Effect because belief-intentionality is relevant for moral judgments about bad outcomes and aim-intentionality is relevant for moral judgments about good outcomes: “[A]gents are generally held to merit moral praise or credit for bringing about good consequences only if they do so aim-intentionally”); Knobe & Burra, *supra* note 1, at 121 (discussing an alternative explanation for the Knobe Effect: that for every intentional action, there must be something that the agent had an intention to do, but to have an intention to cause a specific outcome, the agent must *mean* for it to happen); *accord* Solan, *supra* note 39, at 526 (citing a study by Guglielmo and Malle that used the same scenario involving the CEO and the environment but asked different questions and found that only 10% of respondents said the CEO “intentionally harmed” the environment, but 70% said the CEO “intentionally adopted a profit-raising program that he knew would harm the environment,” and that the respondents assigned blame to both groups (citation omitted)).

45. Solan, *supra* note 39, at 522.

46. *See infra* Part II.C (discussing the hypothesis that the state must condemn antisocial values).

47. Solan, *supra* note 39, at 524. In addition, subjective intentions are not binary but operate on a continuum. *See* Knobe, *supra* note 5, at 327 (arguing that “people’s representation of the agent’s attitude is best understood, not in terms of a simple dichotomy between ‘in favor’ and ‘not in favor,’ but rather, in terms of a whole *continuum* of different attitudes an agent might hold [P]eople can

intentions do not indicate intentional action unless they relate to the morally bad outcome. Sub-baseline intentions about other things—for example, if the CEO intended to steal her neighbor’s car after work—do not indicate that she intentionally harmed the environment.⁴⁸

Solan’s baseline theory finds support in Knobe’s theory of alternative possibilities, which are the morally good actions that people think others should perform and the morally good intentions they think others should have.⁴⁹ When a person causes a bad outcome, people judge her in relation to the good actions and intentions that they would have expected.⁵⁰ Thus, moral judgments about what should have happened influence how people “make sense” of what actually happened—including whether an action was intentional.⁵¹

represent the agent as strongly opposed, as strongly in favor, or as occupying any of the various positions in between.”). This notion comports with the Model Penal Code’s gradations of intent. *See* MODEL PENAL CODE §§ 210.0–210.4 (1985).

48. *See*, Knobe, *supra* note 5, at 322–23 (citing a study by Keys and Pizarro that found that even when one person is made to look like a “generally nice person” and another is made to look like a “generally nasty person,” the Knobe Effect tracks the morality of the action, not the person).

49. *Id.* at 326.

50. *Id.*

51. *Id.* (“[M]oral considerations are playing a role in people’s way of thinking about alternative possibilities. Very roughly, people regard certain possibilities as relevant because they take those possibilities to be especially good or right. With these thoughts in mind, we can now offer a new explanation for the impact of moral judgments on people’s intuitions. The basic idea is just that people’s intuitions in all of the domains we have been discussing—causation, doing/allowing, intentional action, and so on—rely on a comparison between the actual world and certain alternative possibilities. Because people’s moral judgments influence the selection of alternative possibilities, these moral judgments end up having a pervasive impact on the way people make sense of human beings and their actions.”). As this passage shows, Knobe’s research may suggest that moral intuitions influence not only how people perceive intent but also how they perceive causation and permission. *See, e.g.*, Knobe, *supra* note 5, at 319 (referencing studies suggesting that people attribute more causation to the actions of people who did not follow generally accepted rules than to those who did); Solan, *supra* note 39, at 530–31 (referencing studies suggesting that people attribute more causation to the actions of people with morally bad mental states than those with morally good mental states). While questions of causation and permission are beyond the scope of this Article, a skewed perception of causation could exacerbate the harshness of felony murder by punishing not only defendants who lacked the subjective intent to kill but also those who did not even cause, at least in the traditional sense, the death. *See also* Leonard Birdsong, *The Felony Murder Doctrine Revisited: A Proposal for Calibrating Punishment That Reaffirms the Sanctity of Human Life of Co-Felons Who Are Victims*, 33 OHIO N.U. L. REV. 497, 499 (2007) (arguing that felony murder should require proof of proximate cause).

Solan's baseline theory and Knobe's alternative possibilities support the conclusion that felony murder is intentional when the underlying conduct contradicts the presumption that people intend to avoid killing other people. For example, a person falls below that baseline when her conduct demonstrates a disdain for the value of human life. When her conduct is also felonious and causes a death, people judge her even more harshly against the alternative possibilities. Thus, in cases of felony murder, sub-baseline subjective intentions and worse-than-expected conduct combine with the extremely bad outcome of death to give people a reason to say something novel about the defendant's intent. As with the CEO, they say the killing was intentional because the conduct, outcome, and mental state were sufficiently bad.⁵²

Knobe and Solan were not the first to put forward an alternative theory of intent explaining felony murder. For example, James Tomkovicz offered this hypothesis in 1994:

There may well be a difference . . . in how those who study the criminal law and those who elect lawmakers define accidents. Both groups agree that an innocent driver whose vehicle malfunctions in a way that was wholly unforeseeable has killed accidentally. Scholars, however, would say the same of an individual engaged in a felonious enterprise. Because the felon was not negligent, his or her killing was an "accident" by definition. On that point, the public probably disagrees. "Accidental" means innocent, and "innocent" means without fault. The public does not perceive a nonnegligent killing during a felonious endeavor to be lacking in fault. A person who engages in a criminal and likely quite immoral act is not "innocent."⁵³

52. One may ask at this point what distinguishes my version of felony murder from depraved heart murder, which punishes accidental homicide during the course of reckless conduct. I suppose it would be fair to say that my version of felony murder is a subset of depraved heart murder, in that it simply describes depraved heart murders that occur during a dangerous felony. The requirement of an underlying felony is important, however, because it ensures that the underlying conduct, one of Knobe's three parts of intent, was highly immoral. One presumes that only the most immoral conduct would cause people to make the highly unusual statement that one person intentionally killed another, although I know of no research comparing the badness of the conduct with the tendency to call the action intentional.

53. Tomkovicz, *supra* note 12, at 1472 (citation omitted). Of course, the fact that an action was not accidental does not necessarily mean it was intentional. However, accidental actions are never intended. Thus, lawmakers' refusal to view felonious deaths as accidental makes it easier to call them intentional.

As this passage demonstrates, a decade before Knobe's study, scholars had begun to suggest that popular notions of culpability and intent were playing a role in felony murder.⁵⁴ Although Knobe employs new empirical research to support the claim, the idea is not new.

In fact, Knobe's theory parallels the centuries-old concept of *malice aforethought*, a traditional element of murder. In the thirteenth century, *malice aforethought* did not refer to a subjective intent to kill or recklessly endanger but "appears to have meant nothing more definite than a general intention to commit a wrong."⁵⁵ As with felony murder, early conceptions of *malice aforethought* "clearly included felony killings which were later explained by the constructive malice doctrine, the forerunner of felony murder."⁵⁶ It thus appears that the law of homicide has always made room for the notion that a person is guilty of intentionally killing when, instead of subjectively intending to kill, a dangerous felony causes a death.

On the other hand, modern courts and legislatures have abandoned the generalized notion of malice in favor of narrowly defined gradations of intent. If the Knobe Effect simply reflects outdated sensibilities, then perhaps it is not relevant to contemporary murder doctrines. Modern criminal law is explicitly based upon gradations of *subjective* intent,⁵⁷ and by moving away from subjective intent, a conception of felony murder that relies on the Knobe Effect could appear archaic or backwards.⁵⁸ On the other

54. *Id.*

55. Comment, *Felony Murder as a First Degree Offense: An Anachronism Retained*, 66 YALE L.J. 427, 430 (1957) (citations omitted).

56. *Id.* (citations omitted) (describing Herbert's Case as the first ever felony murder). The "constructive malice" theory holds that certain underlying felonies, regardless of the defendant's mental state, make the defendant so culpable that the state is justified in *imputing* the mens rea for murder to her. In other words, it acts as a "mens rea-imposing mechanism," treating the defendant *as though* she had the subjective intent required for murder. See *supra* note 19 for further discussion.

57. See, e.g., MODEL PENAL CODE §§ 210.0–210.4 (1985).

58. In fact, the commenter in the *Yale Law Journal* who noted the connection to thirteenth-century malice ultimately concluded that first-degree felony murder no longer makes sense because first-degree murder requires a subjective intent to kill. Comment, *supra* note 55, at 430–32. This notion was first expressed after the Revolutionary War by the Pennsylvania Legislature, which defined first-degree murder as willful, premeditated, and deliberate. *Id.* at 432. Second-degree murder, on the other hand, may be reckless or negligent, which means that felony murder must necessarily be in the second degree. *Id.* at 431–32. This is really just an argument for the abolition of first-degree felony murder. Of course, felony murder laws are not based solely on the theory of culpability embodied in Solan's baseline but also find support in utilitarian theories and other theories of punishment. See *also infra* Part III.C (discussing the possibility that the Knobe Effect is an

hand, some argue that modern criminal law's exclusive reliance on subjective intent fails to adequately account for other indicators of culpability, especially in felony murder cases.⁵⁹ Treating some felony murders as intentional, then, could reconcile the two positions because both agree that intent demonstrates culpability and culpability justifies punishment.

B. Retributive Theory: Intent Justifies Murder-Level Punishment

When a killing is intentional, retributivists should support imposing murder-level punishment because the retributive theory bases punishment on culpability and culpability on intent.⁶⁰ To be sure, some retributivist scholars have argued that intent, for purposes of determining culpability, refers only to the defendant's mental state.⁶¹ As David Crump observes, these scholars believe that "mens rea is the only legitimate determinant of blameworthiness."⁶² Indeed, even when murder statutes do not require a subjective intent to kill—for example, when murder is based on "gross recklessness . . . of a sufficient magnitude to evince a callous or depraved indifference to the value of human life"⁶³—culpability is still determined by the defendant's mental state. Nevertheless, retributivist scholars, if dedicated to the idea that intent determines culpability, should revise their view of culpability in light of Knobe's research. If Knobe's

undesirable part of our psychology that the law should push us to abandon in favor of more progressive and rational ways of thinking).

59. See generally Crump, *supra* note 12; Guyora Binder, *The Culpability of Felony Murder*, 83 NOTRE DAME L. REV. 965, 975–76 (2008).

60. See Binder, *supra* note 59; see also *id.* at 1001 ("[Jeremy] Bentham's distinction between motive and intent combines three ideas: (1) the criminal law should reduce discretion by precisely defining offenses; (2) it should define offenses in neutral descriptive language rather than normative language; and (3) it should define culpability by reference to cognitive states like expectations, rather than desiderative states like purposes." (citing JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 92 (J.H. Burns & H.L.A. Hart eds., Univ. of London 1970) (1789))).

61. See, e.g., Tomkovicz, *supra* note 12, at 1439–40 (arguing that the law should impose a "gradation [of punishments] proportionate to the established level of mental fault" (emphasis added)); Binder, *supra* note 59, at 989 n.109 (citing Larry Alexander, *Reconsidering the Relationship Among Voluntary Acts, Strict Liability, and Negligence in Criminal Law*, 7 SOC. PHIL. & POL'Y 84, 101 (1990); Jerome Hall, *Negligent Behavior Should Be Excluded from Penal Liability*, 63 COLUM. L. REV. 632, 634–37 (1963)).

62. Crump, *supra* note 12, at 1159.

63. Tomkovicz, *supra* note 12, at 1439–40.

theory has truly deepened the meaning of intent,⁶⁴ then a defendant's culpability should be judged by all aspects of such intent, including conduct and outcomes.⁶⁵

Some retributivist scholars, in fact, argued for a broader definition of culpability before the Knobe Effect was discovered. David Crump and Guyora Binder, for example, argued that a person's culpability depends on essentially the same factors that Knobe would use to describe intent.⁶⁶ Their position is that conduct and outcomes demonstrate a person's culpability,⁶⁷ which is consistent with Knobe's theory that conduct and outcomes demonstrate intent.⁶⁸ Importantly, by describing the connection between conduct, outcomes, and culpability, Crump and Binder unknowingly sketched a framework for restricting felony murder to killings that would be labeled intentional under Knobe's theory.

Crump elucidates the pre-Knobe position that culpability depends on conduct and outcomes:

[T]he drafters [of the Model Penal Code] seem to be saying that mens rea is the only legitimate determinant of blameworthiness, [but] the argument rests upon debatable propositions. . . . The criminal law has never been limited to mens rea alone in assessing the severity of crime. Actus reus and results count, too.⁶⁹

Binder offers a more detailed account of how conduct and outcomes indicate culpability. He argues that a person's conduct "expresses" her values and the resulting outcomes "gratify" them.⁷⁰ Criminal punishments must therefore address the "expression" and "gratification" of bad values in the same way that they address

64. See *infra* Part III.C (discussing whether the Knobe Effect simply represents a mistaken perception of mental states).

65. It is important to note here that all retributivist laws depend on perceptions of culpability that are fraught with uncertainty and bias. For example, hindsight bias and outcome bias make people more likely to say that a traditional murder defendant subjectively intended to kill or recklessly endanger. See *infra* Part III.C. Therefore, even though a felony murder defendant's perceived culpability may be fraught with error, such inaccuracies affect traditional murder convictions as well. Indeed, perhaps they affect traditional murder more than felony murder because the only way to determine intent in a traditional murder case is through the impossible task of determining what was going on in the defendant's mind.

66. See Crump, *supra* note 12, at 1159–60, 1162; Binder, *supra* note 59, at 974–75.

67. See Crump, *supra* note 12, at 1159–60, 1162; Binder, *supra* note 59, at 974–75.

68. Knobe & Burra, *supra* note 1, at 123–25.

69. Crump, *supra* note 12, at 1159–60, 1162 (citations omitted).

70. Binder, *supra* note 59, at 1031–33.

culpable mental states. Binder calls this the “expressive theory” of culpability.⁷¹

Binder rejects the notion that a person’s culpability depends on her subjective mental state alone—what he calls the “cognitive view” of culpability:⁷²

A second reason for the persistent view of felony murder as rationally indefensible is the narrowly cognitive view of culpability that prevails in contemporary criminal law theory. According to this view, culpability is purely a function of the expectation of harm attributable to an actor at the time he or she acts. Thus, the actor’s purposes, motives, meanings, and values are irrelevant. In particular, such goals as completing a rape, demeaning a victim because of her race, or intimidating political opponents are irrelevant to culpability for a killing. All that matters is the death and the expectation of causing it. This cognitive conception of culpability reflects a restrictive view of the role of criminal law in a liberal state—as opposing harmful conduct, but taking no sides in disagreements about values.⁷³

Binder argues that a defendant’s values are an important indicator of culpability because the state’s social contract with its citizens requires punishing the expression of bad values. His idea is based on John Rawls’s theory that the purpose of the state is to “secure to each individual the broadest sphere of freedom compatible with like freedom for others.”⁷⁴ For the state to operate well, it must ensure that citizens embody the kinds of values that enhance freedom for everyone. The state therefore cannot remain neutral toward people who express values contrary to the Rawlsian social contract. When people express antisocial values through conduct that causes harm, the state must condemn and punish those values.⁷⁵

71. *Id.* at 974–75.

72. *Id.* at 970.

73. *Id.* David Crump agrees with Binder, maintaining that blameworthiness depends on more than subjective intent. Crump, *supra* note 12, at 1159 (criticizing the Model Penal Code because “the drafters seem to be saying that mens rea is the only legitimate determinant of blameworthiness, that the traditional determinants of mens rea for murder are the only way to describe the appropriate mental states for murder Again, the argument rests upon debatable propositions”).

74. Binder, *supra* note 59, at 1018–19.

75. *Id.* at 974–75, 1025, 1032. Binder further argues that the Model Penal Code’s causation standard reflects the intuition that bad values should be punished. *Id.* at 1025, 1032 (“The Model Penal Code causation standard restates the scope of the risk problem in explicitly normative terms. It asks whether ‘the actual result differs from the probable [or designed] result only in the respect that a

Binder posits that even unexpected outcomes affect how much punishment is deserved. If the person's conduct expresses bad values, he says, then a bad outcome strengthens the expression: the worse the outcome, the stronger the expression of bad values, and thus more punishment is deserved.⁷⁶ In this way, the morality of a defendant's conduct *and* the morality of the outcomes it causes determine culpability. When a defendant's bad conduct causes a death, greater punishment is deserved even if the conduct and mental state are unchanged.⁷⁷

Binder and Crump agree, therefore, that a properly limited version of felony murder would be justified under retributive

different person . . . is injured . . . or . . . the actual result involves the same kind of injury . . . as the probable [or designed] result and is not too remote or accidental in its occurrence to have a just bearing on the actor's liability.' What is the 'same kind of injury'? When is the 'only' relevant difference between two injuries the identity of the victim? When is a result not 'too remote or accidental' to justly affect liability? These are obviously normative questions depending in part on our evaluations of the aims of the actors. . . . [Instead, on my] view, action expresses value by identifying us with normative social practices. In *The Morality of Freedom*, the legal philosopher Joseph Raz denies that our desires determine our goals, and argues instead that our desires often flow from normative beliefs about what is best for us. Thus, he contends, we act on the basis of normative reasons or *values*, rather than unreflective wants. In *Value in Ethics and Economics*, the moral philosopher Elizabeth Anderson offers an institutional account of value as a social practice of recognizing certain kinds of goods" (first and second alterations in original) (citations omitted) (citing JOSEPH RAZ, *THE MORALITY OF FREEDOM* 288-320 (1986); ELIZABETH ANDERSON, *VALUE IN ETHICS AND ECONOMICS* 6-7, 11-15 (1993)).

76. *Id.* at 974. For example, Binder argues that a successful murderer deserves more punishment than an attempted murderer, even though their conduct expressed the same bad values. *Id.* at 1028.

77. *Id.* at 1037. On the other hand, some say it is irrational to increase punishment based on bad outcomes. *See, e.g.*, Tomkovicz, *supra* note 12, at 1471 & n.169 (arguing that an "injury to another person is a weighty factor in the balances struck by the public. The more serious the injury is, the weightier the factor is. There is no need to pay a person back or to make a person pay merely because of the damage done, but damage makes us begin to think along those lines and generates an inclination to respond. . . . Professor Schulhofer refers to [this response] as 'retaliation.' Others have called it 'expiation' and 'vengeance.' Whatever its name, the public possesses a certain attachment to the concept. That attachment is part of the substructure of the felony-murder rule." (citing Stephen J. Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497, 1571 (1974); David Crump & Susan W. Crump, *In Defense of the Felony Murder Doctrine*, 8 HARV. J.L. & PUB. POL'Y 359, 368 (1985); Note, *Felony Murder: A Tort Law Reconceptualization*, 99 HARV. L. REV. 1918, 1932 (1986)); *see also infra* Part III.C.3 (discussing whether the Knobe Effect is simply evidence of a predisposition to blame people who cause bad outcomes, regardless of whether they are culpable).

theory.⁷⁸ Binder argues from his expressive theory, contending that a defendant who causes a death in the course of an inherently dangerous felony deserves to be punished as a murderer because the dangerous conduct, combined with the outcome of death, expresses antisocial values as strongly as, if not more strongly than, traditional murder.⁷⁹ Crump similarly argues that felony murder is a necessary part of the criminal law because traditional murder, which cannot impose murder-level punishment without proof of subjective intent, does not account for the full range of factors that determine culpability, including conduct and outcomes.⁸⁰ Crump thus argues that felony murder “may actually serve the policy of linking the criminal law to moral blameworthiness.”⁸¹

Building upon Crump and Binder, I propose that the best way to restrict felony murder to defendants who are as culpable as traditional murderers is to confine it to cases in which the killing would be called intentional under Knobe’s theory. Felony murder would thus fill a gap in homicide schemes, addressing the set of intentional killings not covered by traditional murder while also eliminating instances in which less culpable defendants are punished too harshly. If such a rule is desirable, then the next question is how to structure it to apply only to intentional killings.

C. Restricting Felony Murder to Intentional Killings

Although Knobe’s theory suggests that felony murder sometimes involves intentional killings, certainly not all deaths that occur during the commission of a felony are intentional. As

78. See generally Crump, *supra* note 12; Binder, *supra* note 59.

79. See generally Binder, *supra* note 59.

80. See generally Crump, *supra* note 12.

81. *Id.* at 1162. Moreover, culpability should perhaps not depend exclusively on the actor’s subjective intent, especially in light of studies showing that perceptions of others’ mental states are inherently flawed. See, e.g., Christopher Slobogin, *The Civilization of the Criminal Law*, 58 VAND. L. REV. 121, 146 (2005) (“[E]xperience and research demonstrate that judicial and jury conclusions about core culpability concepts such as premeditation, recklessness, and insanity differ significantly across individuals and across juries. Given this unreliability, many of these conclusions about blameworthiness cannot possibly achieve the 90 to 95 percent degree of accuracy normally associated with the reasonable doubt standard. That should not be surprising, given the ill-defined scope of legal mental states and the difficulty of investigating subjective beliefs and desires. But it is disturbing because these unreliable assessments can spell the difference between a conviction for manslaughter and eligibility for the death penalty, or between a prison term and indeterminate institutionalization in a mental hospital.” (citations omitted)); see also *infra* Part III.C (discussing the possibility that the Knobe Effect does not affect the meaning of intent but is merely evidence of the inability to perceive others’ subjective intentions).

described above, a killing cannot be considered intentional under Knobe's theory unless the killer exhibits sub-baseline intentions regarding the death. Unrelated intentions, even if morally bad, are not relevant.⁸² Thus, purely unforeseeable or accidental killings should not be considered felony murder because they exhibit no sub-baseline intentions regarding the death. In this respect, Crump and Binder are instructive once again. By devising ways to limit felony murder to actions that demonstrate murder-level culpability, they unknowingly provide a framework for limiting the rule to killings that would be considered intentional under Knobe's theory.

Binder would limit felony murder to cases involving inherently dangerous felonies.⁸³ Under his "expressive theory," a felony murder conviction is not justified unless the defendant harbored bad values that were "indulged" or "gratified" by the death.⁸⁴ He therefore argues that only inherently dangerous felonies, such as those in which the defendant "uses fatal violence to coerce a victim's cooperation or overcome her resistance during a crime like rape or robbery," express and gratify such bad values.⁸⁵

Most states appear to agree with Binder on this point. As of 2005, 35 states had restricted felony murder to a list of felonies that pose the greatest risk to life.⁸⁶ When a death results from one of

82. See *infra* note 100 and accompanying text; see also Knobe, *supra* note 5, at 323 (describing a 2009 study suggesting that the perception of intent is tied to the morality of conduct, not blame for the resultant outcome).

83. Binder, *supra* note 59, at 988. This limitation makes sense considering the way in which he connects culpability to outcomes:

If we punished attempts and completed crimes equally, successful offenders would be left more satisfied than unsuccessful attempters. Their regret at having been caught and punished would be mitigated by their pleasure in having achieved their criminal aims. From this viewpoint, we are obliged to punish the successful wrongdoer more than the attempter lest we become complicit in his self-indulgence by permitting his undeserved gratification. This is, in my view, a very strong argument that punishing harm is deserved.

Id. at 1028 (citation omitted).

84. *Id.* at 1027–29, 1039–40.

85. *Id.* at 1037, 1045–46. An obvious counterargument is that, even if a defendant decided to expose someone to a risk of death to perpetrate a rape or robbery, the lack of subjective intent to kill means that the death did not gratify the defendant's values, at least not to the extent that Binder suggests.

86. Birdsong, *supra* note 10, at 21 & n.140. See also Tomkovicz, *supra* note 12, at 1467; Crump, *supra* note 12. Birdsong and Crump agree with this view of felony murder, arguing that the sanctity of human life is the most important value underlying the felony murder rule. Birdsong, *supra* note 51, at 499, 506 ("[S]tates might wish to choose to amend their felony murder statutes in an effort to effectuate the proximate cause theory of liability. Such an approach could prove to be a better way to calibrate punishment in a proportionate way that would value the lives of co-felons who are victims. Such calibration would better serve to deter

those felonies, it is easier to say that the defendant expressed disdain for human life.⁸⁷ Furthermore, as Crump observes, the fact that a majority of states limits felony murder to inherently dangerous felonies shows that the version of the rule commonly criticized by scholars—the version that punishes “any felony that results however unpredictably in a death”—is not representative of the general rule.⁸⁸

Crump, similar to Binder, would limit felony murder to crimes committed in a manner “clearly dangerous to human life.”⁸⁹ This formulation attempts to restrict murder-level punishment to cases in which the defendant’s conduct expressed a high level of disdain for human life.⁹⁰ The difference is that Crump does not use a

felons who might kill while also reaffirming the sanctity of human life, even the life of co-felons who may be killed. . . . Under the proximate cause theory, liability attaches for any death proximately resulting from the unlawful activity, though it is most often applied in those states that follow the enumerated felonies that are inherently dangerous to life approach to felony murder.” (citation omitted)). See also Crump, *supra* note 12, at 1163 (“The rule also performs a function involving condemnation, because it reaffirms the sanctity of human life by reserving severe sanctions for crimes that destroy human life.” (citation omitted)). Though the sanctity of human life may not be essential to the Rawlsian social contract, it is nevertheless important for the law to uphold. Maximizing freedom and social good for all citizens, moreover, arguably depends on everyone’s right to remain alive.

87. Indeed, the earliest forms of felony murder did not encompass accidental killings. Binder, *supra* note 20, at 202 (“Apart from one case predicated on a bungled suicide and four cases predicated on putatively consensual abortions, none of the known felony murders punished in nineteenth-century America could plausibly be described as accidental. In almost all of these cases, death resulted from the deliberate infliction of violence.”). See also Karen M. Quinn, Case Note, *Criminal Law—A Reckless Indifference to Human Life Is Sufficient Evidence to Prove Culpability in a Felony-Murder Case and Therefore Imposition of the Death Penalty Is Not a Violation of the Eighth Amendment—Tison v. Arizona*, 107 S. Ct. 1676 (1987), 37 DRAKE L. REV. 767 (1988). Quinn points out that reckless indifference is enough to justify the death penalty under the Eighth Amendment ruling in *Tison v. Arizona*. *Id.* at 776 (citing *Tison v. Arizona*, 107 S. Ct. 1676, 1684–68 (1987) (holding that the possibility of killing someone is inherent in the commission of any violent felony and thus one can say that all violent felons constructively foresee a killing)). Donald Baier thus argues that *Tison* makes the most sense when felony murder is limited to inherently dangerous felonies or felonies committed in clearly dangerous ways, which is consistent with Crump and Binder. Donald Baier, Note, *Arizona Felony Murder: Let the Punishment Fit the Crime*, 36 ARIZ. L. REV. 701, 705–07 (1994).

88. Crump, *supra* note 12, at 1176–77. Of course, individual judges and prosecutors may ignore such limiting language.

89. *Id.* at 1171.

90. *Id.* at 1166 (arguing that restricting felony murder to crimes committed in ways that are clearly dangerous to human life “ties the crime of murder to

predetermined list of “inherently dangerous” felonies because even those felonies are sometimes committed in ways not clearly dangerous to human life.⁹¹ Crump’s formulation thus attempts to tailor felony murder to cases in which the defendant *actually* expressed disdain for human life.⁹²

Perhaps Crump’s formulation finds the most support in Knobe’s theory. One can be relatively certain that when a defendant commits a felony in a way that is “clearly dangerous to human life,” her subjective intentions regarding the risk of death fall below Solan’s baseline. Simply, she does not care enough about endangering others’ lives. Thus, when she ends up killing a person in the course of the felony, Knobe’s theory would label the killing intentional.⁹³

Binder’s formulation, on the other hand, is less precise because there is no felony that endangers human life *every* time it is committed.⁹⁴ Some felonies tend to be more dangerous than others, and those are typically included on states’ lists of “inherently” dangerous felonies, but there are always some cases in which those crimes—even armed robbery or violent rape—are committed without actually endangering human life.⁹⁵ Even so, merely deciding to commit a felony that carries a high risk of being

relatively high degrees of individual blameworthiness”). *But see* Tamu Sudduth, Comment, *The Dillon Dilemma: Finding Proportionate Felony-Murder Punishments*, 72 CAL. L. REV. 1299, 1320–22, 1326 (1984) (arguing that the felony murder rule would be unnecessary if all homicides were punished in proportion to the degree of disdain for human life expressed by the crime: “If the totality of the circumstances surrounding a homicide indicates that the defendant consciously subjected a human life to unreasonable risk, then murder-level punishment is justified. If the defendant’s conduct does not suggest a recklessness or an intent to kill, then the defendant’s culpability is insufficient to invoke murder penalties.”).

91. Crump, *supra* note 12, at 1166, 1171 (criticizing California’s felony murder law, which is limited to inherently dangerous felonies in the abstract: “The principal limitation upon the felony murder rule in California is the ‘inherently dangerous felony’ requirement. This concept differs sharply from the ‘clearly dangerous act’ requirement in the state statute discussed above. In California, the relevant question is whether the felony ‘in the abstract’ is inherently dangerous. This formulation is subject to criticism because it divorces the definition of murder from the individual blameworthiness of the defendant. . . . But that is not all. The California court has had a great deal of trouble deciding precisely which felonies are ‘dangerous.’” (citations omitted)).

92. *Id.* at 1166.

93. But see *supra* Part II.B for a discussion of whether legal culpability has traditionally been determined by intent qua mental state or as defined more broadly by Knobe.

94. See Crump, *supra* note 12, at 1166, 1175.

95. *Id.*

dangerous to human life likely reveals subjective intentions below Solan's baseline.⁹⁶

Perhaps lawmakers could combine the theories of Binder and Crump to ensure that felony murder covers only intentional killings. Restricting felony murder to situations in which both conditions are satisfied—in which the crime is on the list of inherently dangerous felonies and was committed in a manner clearly dangerous to human life—would best ensure that the defendant had the kinds of sub-baseline intentions that, along with morally bad conduct and outcomes, indicate that the killing was intentional. In this way, properly restricted felony murder laws could single out defendants deserving of murder-level punishment and thus find support in retributive theory.⁹⁷

96. For a list of state laws that restrict felony murder to a predetermined set of crimes typically dangerous to human life, see Birdsong, *supra* note 10, at 21 n.140. Of course, one can question whether the crimes included on the list are, in fact, dangerous enough to merit inclusion. Additionally, Binder would modify the felony murder rule based on the values of each jurisdiction. See Binder, *supra* note 20, at 201–05. That is, the extent to which a certain value is “antisocial” and deserving of punishment depends on the particular values of the locale. *Id.* A related issue in this context is how felony murder might vary across jurisdictions with different values. Could a homicide be intentional in one jurisdiction but not another? Perhaps this problem suggests a need for further research regarding how the Knobe Effect operates across different cultures and demographics. Such research is in its infancy. One study suggests that the Knobe Effect operates in the Hindi language but not necessarily in India itself (the subjects of the study were in the United States). Knobe & Burra, *supra* note 1, at 126–29. Another study, unpublished, suggests that the Knobe Effect may be weaker among Asian-American males. See Lawrence Ngo, Walter Sinnott-Armstrong & Scott Hueteel, Individual Differences in the Knobe Effect and Experimental Philosophy (2011) (unpublished study). Moreover, it would appear difficult to discern whether the relevant subset is geographic or demographic (or both). Binder would choose geography, arguing that the relevant values are those expressed by the *law* of each jurisdiction. Binder, *supra* note 20, at 207 (“One function of a felony murder rule, then, is to work in conjunction with other rules of criminal liability to map a particular society’s moral intuitions about violence and malice. This means that there can be no universally valid answer to the question of the justice of ‘the’ felony murder rule. Instead, we must evaluate each felony murder rule as it is defined and put into practice in a particular jurisdiction, in a particular legal and cultural context.”). See also Daniel J. Abbott, *A Comparative Analysis of Felony Murder in Developed and Under-Developed Nations* (Am. Soc’y of Criminology, Annual Meeting Presentation, Oct. 30–Nov. 2, 1975).

97. This restriction also appears to track the traditional definition of *malice aforethought*. See Tomkovicz, *supra* note 12, at 1439 n.38 (“The traditional and still common definition of murder is a killing with malice aforethought. The essence of malice is this ‘callous or depraved indifference to human life.’ Intentional or knowing killings qualify as murders because they satisfy this minimum essence. In fact, they entail *even more* culpability than the minimum indifference. Grossly reckless killings also satisfy this minimum threshold for malice.” (citations omitted)).

IV. COUNTERARGUMENTS

Although some felony murders would be considered intentional under Knobe's theory, limiting the rule to such cases would not necessarily satisfy retributivist critics. As explained by Tomkovicz and Judge Rudolph Gerber, felony murder suffers from particular deficiencies that may not be remedied by restricting it to intentional killings. Additionally, Knobe's theory may suggest that a killing was not intentional unless the killer *foresaw* or *desired* the death, which would drastically reduce the percentage of intentional felony murders. Finally, some researchers argue that the Knobe Effect has no bearing on the meaning of intent and simply represents a mistaken perception of mental states. If this is true, then felony murders are never intentional, and Knobe's theory cannot answer retributivist critics at all.

A. Felony Murder Is Especially Corrupt and Unfair

If felony murder operates in especially unfair or corrupt ways, then it may not be justifiable even when restricted to intentional killings. Rudolph Gerber, a judge on the Arizona Supreme Court, argues that the felony murder rule is corrupted by politics and unfair to individual defendants.⁹⁸ He maintains that while society may feel a need to condemn antisocial values (as Binder would argue), it is fundamentally unfair to heap society's hatred for crime onto a few unfortunate individuals.⁹⁹ The nature of the political process exacerbates the problem, he says, because the defendant's status as a felon encourages lawmakers and judges to impose excessive punishment: "By objectifying the [felon] as nothing more than an undeserving criminal, lawmakers find it easier to aggravate punishment even to death, no matter the price of caprice."¹⁰⁰

Tomkovicz levies a similar charge, arguing that felony murder's unfairness stems from the political process.¹⁰¹ Legislatures and

98. See generally Rudolph J. Gerber, *The Felony Murder Rule: Conundrum Without Principle*, 31 ARIZ. ST. L.J. 763 (1999). Judge Gerber does not mention Binder's expressive theory, but his arguments are directly relevant. See *id.*

99. *Id.* at 782-84.

100. *Id.* at 783 (internal quotation marks omitted) (citing Note, *Felony Murder: A Tort Law Reconceptualization*, 99 HARV. L. REV. 1918, 1931 (1986)). Gerber's argument could be seen as a criticism of Solan's baseline theory. As Solan hypothesized, people would call a homicide intentional if it involved conduct exhibiting sub-baseline intentions regarding the possibility of death. Gerber's response would be that it is unfair to apply different rules to people who fall below the baseline, especially considering that felony defendants are "the very ones who need to learn something about principle." *Id.* at 782-84.

101. Tomkovicz, *supra* note 12, at 1461-63.

judges, he says, feel pressure to placate a populace that does not care about treating felons fairly.¹⁰² This pressure often overrides their ability to maintain “consistency and fairness” when charging and punishing felony murder defendants:¹⁰³

The demand for “law and order” strikes an emotional chord in America. One can hardly be elected to public office without embracing the concept wholeheartedly. . . . [A]nyone bent on reforming the [felony murder] rule must fight the tide and be prepared to pay a political price. In the world of American politics, logical consistency and fairness to felons are not very potent weapons against the charge that one is soft on crime and hostile to law and order. . . . [P]robably inextricable from the public consciousness is the idea that felons—by virtue of their choices to engage in felonies—have effectively forfeited any entitlement to close scrutiny of their blameworthiness.¹⁰⁴

While these complaints may ring true in jurisdictions that treat purely accidental killings as murder,¹⁰⁵ Knobe’s theory, again, does not support such an expansive rule. Knobe’s theory requires that the defendant demonstrate sub-baseline intentions regarding the victim’s life before the killing can be called intentional.¹⁰⁶ The “accidental death” version of felony murder thus finds no support in Knobe’s theory because purely accidental killings should not be called intentional.

Moreover, the principal case cited by Judge Gerber, *People v. Aaron*,¹⁰⁷ actually supports the limited versions of felony murder suggested by Crump and Binder. While *Aaron* technically abrogated felony murder in Michigan, it expressly sanctioned murder-level punishment for defendants who unintentionally killed while committing felonies that exhibited disdain for human life.¹⁰⁸ The

102. *Id.*

103. *Id.*

104. *Id.* at 1461–63, 1475 (citations omitted).

105. See Gerber, *supra* note 98, at 763, 765, 767, 771, 778, 782–83; Tomkovicz, *supra* note 12, at 1438, 1444, 1449, 1472–73.

106. See *supra* Part II.C.

107. Gerber, *supra* note 98, at 784 (citing *People v. Aaron*, 299 N.W.2d 304, 327 (Mich. 1980)).

108. *People v. Aaron*, 299 N.W.2d 304, 327 (Mich. 1980) (noting that “in many circumstances the commission of a felony, particularly one involving violence or the use of force, will indicate an intention to kill, an intention to cause great bodily harm, or wanton or willful disregard of the likelihood that the natural tendency of defendant’s behavior is to cause death or great bodily harm. Thus, the felony-murder rule is not necessary to establish mens rea in these cases”). See also Crump, *supra* note 12, at 1159–60, 1164.

court observed that in such cases, felony murder is irrelevant because the state can prove the intent required for murder.¹⁰⁹ Knobe's theory, in accordance with 35 states, says essentially the same thing: A killing is not intentional unless the conduct and outcomes are morally bad (as in all cases of felony murder) *and* the defendant exhibits sub-baseline intentions regarding the death (as in the formulations suggested by Crump and Binder). It is thus mostly a question of semantics whether *Aaron* "abrogated" felony murder or simply changed it to the version suggested by Crump and Binder.¹¹⁰ In this way, Knobe's theory suggests that felony murder is not necessarily "unfair" or "capricious" but is consistent, when properly restricted, with the meaning of intent.

B. Knobe's Theory Does Not Apply to Unforeseen Outcomes

According to some experts, the Knobe Effect occurs only when the actor *knows* that she will cause the relevant outcome.¹¹¹ If this is true, then Knobe's theory may be irrelevant to most felony murders because, in most cases, the defendant did not know that her conduct would cause a death.¹¹²

Al Mele and Paul Moser argue that when a person does not believe she is in control of causing an outcome, it makes no sense to say she intended it.¹¹³ Thomas Nadelhoffer explains their argument in this way:

On [their] view, skill and control are necessary conditions of our everyday concept of intentional action [For example, in a scenario in which Lisa and Mike roll a winning number with a pair of dice], Mele and Moser claim that . . . neither Lisa nor Mike performed their respective actions intentionally because in both cases the actions were not the result of any relevant skill on their part—i.e., their actions were simply the fortunate result of chance or luck.¹¹⁴

Frank Hindriks agrees with Mele and Moser, arguing that the Knobe Effect occurs only when the person foresees the morally

109. *Aaron*, 299 N.W.2d at 327.

110. *Id.*; Crump, *supra* note 12, at 1159–60, 1164.

111. *See generally* Knobe, *supra* note 5 (summarizing various studies and debates about the Knobe Effect, all of which deal with scenarios in which the actor foresaw the outcome).

112. I know of no study examining the percentage of felony murder cases in which the defendant knew that her action would cause a death, but I presume it is small.

113. Nadelhoffer, *supra* note 4, at 341, 344.

114. *Id.*

relevant outcome.¹¹⁵ The Knobe Effect, he says, is simply a condemnation of the person's failure to avoid foreseen negative consequences.¹¹⁶

[F]oresight betrays a guilty mind as much as intent does: both reveal that the agent is not properly motivated to avoid an illegal state of affairs. This commonality warrants our judgment that the state is brought about intentionally, even when unintended. In contrast to Knobe, I thus retain the idea that acting intentionally is acting with a certain frame of mind.¹¹⁷

If these researchers are correct, then perhaps Knobe's theory is irrelevant to most felony murder cases. Like Lisa and Mike, most felony murderers do not believe that their conduct will cause the relevant outcome (death). Instead, the death, like the roll of the dice, involves a significant amount of luck. A recent study, however, suggests that the Knobe Effect may actually occur even when the actor does not know or believe that her conduct will cause the relevant outcome.¹¹⁸ If this is true—if the Knobe Effect extends this far—then perhaps Knobe's theory does speak to unforeseen killings.¹¹⁹

The study asked participants to examine four hypothetical scenarios and evaluate whether the person in each acted intentionally.¹²⁰ The scenarios involved a morally good outcome accomplished by skill, a morally good outcome resulting from luck, a morally bad outcome accomplished by skill, and a morally bad outcome resulting from luck.¹²¹ Briefly, the scenarios involved Jake shooting a rifle. The morally good outcome was that he shot a bull's-eye and won a contest. The morally bad outcome was that he shot his aunt to hasten his inheritance. To base the outcome on skill,

115. Knobe, *supra* note 5, at 339.

116. *Id.*

117. *Id.*

118. Nadelhoffer, *supra* note 4, at 341, 345–46. As Nadelhoffer says, “[S]ome simple psychological experiments . . . show that people’s judgments concerning whether actions are intentional can often be affected by the moral features of these actions—features that may trump considerations of skill, luck, and control. Finally, I conclude that if this is correct, philosophers who claim that skill and control are necessary conditions . . . of intentional action appear to be mistaken.” *Id.* at 341.

119. Crump agrees that felony murder should not be restricted to cases in which the defendant subjectively believed that her conduct was likely to cause a death because other kinds of murder do not require such precise subjective intent—for example, “depraved heart” murder. Crump, *supra* note 12, at 1168.

120. Nadelhoffer, *supra* note 4, at 345–46.

121. *Id.*

the hypothetical stipulated that he was an expert marksman. To base it on luck, it stipulated that he had never fired a gun before.¹²²

The remarkable results involve Jake's lucky shots. Even though he did not foresee the consequences of his conduct—having never fired a gun before, he did not believe he could hit his aunt or the bull's-eye—participants in the study called his action intentional more often when the outcome was morally bad.¹²³ Seventy-six percent of participants said he intentionally shot his aunt, while only 28% said he intentionally shot the bull's-eye.¹²⁴ These results suggest that the Knobe Effect occurs even when the actor does not know that her conduct will cause the relevant outcome.¹²⁵

On the other hand, Jake hoped that he would hit the targets and knew that doing so was possible.¹²⁶ Although he did not believe that he would hit either target, he knew that they would be in his line of fire and that he could hit them by chance.¹²⁷ Thus, although the study suggests that the Knobe Effect occurs even when the person does not *foresee* the relevant outcome, it says nothing about situations in which the person has *no idea* that the outcome is

122. *Id.* Nadelhoffer describes the scenarios like this: “(1A): Achievement/Skill: Jake desperately wants to win a rifle contest. He knows that he will only win the contest if he hits the bull’s-eye. He raises the rifle, gets the bull’s-eye in the sights, and presses the trigger. Jake is an expert marksman. His hands are steady. The gun is aimed perfectly. The bullet lands directly on the bull’s-eye. Jake wins the contest. (1B): Achievement/No Skill: Jake desperately wants to win a rifle contest. He knows that he will only win the contest if he hits the bull’s-eye. He raises the rifle, gets the bull’s-eye in the sights, and presses the trigger. But Jake isn’t very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild. Nonetheless, the bullet lands directly on the bull’s-eye. Jake wins the contest. (2A): Immoral/Skill: Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. Jake is an expert marksman. His hands are steady. The gun is aimed perfectly. The bullet hits her directly in the heart. She dies instantly. (2B): Immoral/No Skill: Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. But Jake isn’t very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild. Nonetheless, the bullet hits her directly in the heart. She dies instantly. . . . Each subject was presented with one of these four vignettes along with the following question: ‘Did Jake intentionally kill his aunt [hit the bull’s-eye]?’” *Id.* at 345 (alteration in original) (citation omitted).

123. *Id.* at 345–46.

124. *Id.* at 346. Unsurprisingly, the skillful scenarios also exhibited the Knobe Effect. When Jake was an expert marksman, 95% of respondents said he shot his aunt intentionally, whereas only 79% said he shot the bulls-eye intentionally. *Id.*

125. The percentages in this study were similar to Knobe’s original study, in which participants perceived intent 87% of the time for the bad outcome and 20% of the time for the good outcome. Knobe & Burra, *supra* note 1, at 118.

126. *See* Nadelhoffer, *supra* note 4, at 345

127. *See id.*

possible or does not *desire* it.¹²⁸ If the Knobe Effect does not apply in those situations, then it may not apply in many felony murder cases because most defendants probably did not hope to kill and may not have known that death was even a possibility. As of this writing, no studies have tested whether the Knobe Effect applies in such situations.¹²⁹

C. The Knobe Effect Represents a Mistaken Perception

Some argue that the Knobe Effect represents a mistaken perception of mental states. On this view, morally bad conduct and outcomes cause observers to mistakenly conclude that bad outcomes were subjectively intended.¹³⁰ If the Knobe Effect simply reflects this error, then Knobe's theory cannot provide a rational restriction on felony murder and may actually support abolishing it to the

128. Cf. FRITZ HEIDER, *THE PSYCHOLOGY OF INTERPERSONAL RELATIONS* 83 (1958) (arguing that "trying" and "intending" are equivalent).

129. On the other hand, one might say that committing a felony, or at least a dangerous felony, implies some level of awareness that someone's life is being endangered. See *supra* Part II.C. Additionally, building upon the study involving Jake, Nadelhoffer conducted another study involving a hypothetical actor named Fred, an employee at a nuclear reactor who luckily guesses the numbers of a crucial code. Not surprisingly, when Fred correctly guessed the code to cause a bad result, observers rated his action as very intentional, but when he correctly guessed the code to cause a good result, observers rated his action as less intentional. Nadelhoffer, *supra* note 4, at 347–49. The study also evaluated two other dimensions: the extent to which observers praised or blamed Fred for the result and the distinction between whether he intentionally guessed the right numbers or intentionally caused the ultimate result. *Id.* The results of Nadelhoffer's study are as follows. When Fred guessed the code to prevent an imminent meltdown of the nuclear reactor, the responses were thus:

Q1: 38% said Fred punched in the correct numbers intentionally.

Q2: The average praise rating was 3.0 on a 6-point scale.

Q3: 73% said Fred intentionally prevented the explosion.

Q4: The average praise rating was 4.0 on a 6-point scale. *Id.* at 347.

Yet when he guessed the code to destroy the plant, the results were thus:

Q1: 67% said that Fred intentionally punched in the correct numbers.

Q2: The average blame rating was 5.23 on a 6-point scale.

Q3: 83% said that Fred intentionally caused the explosion.

Q4: The average blame rating was 5.31 on a 6-point scale. *Id.* at 348.

Interestingly, when the moral element of the scenario was removed, the participants in the study were much less likely to call the action intentional. *Id.* at 348–49. That is, when Fred guessed the code to win the lottery, the results were:

Q1: 80% said that Fred did not intentionally punch in the correct numbers.

Q2: 67% said that Fred did not intentionally win the lottery. *Id.* at 349.

130. See, e.g., Knobe, *supra* note 5, at 321–24; Thomas Nadelhoffer, *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, 9 PHIL. EXPLORATIONS 203, 205–06 (2006).

extent it depends on the mistake. There are at least three ways in which the Knobe Effect could represent a mistaken perception of mental states: It may be infected by hindsight bias, motivational bias, and outcome bias. I use the term “infected” because these biases, although they contribute to the Knobe Effect, are likely not its sole cause. Even so, such problems would undermine the argument that felony murder deals with intentional killings.¹³¹

1. Hindsight Bias

Hindsight bias, perhaps the simplest of the three biases, may infect both Knobe’s theory and felony murder. Generally speaking, hindsight bias is “the phenomenon that people overestimate the predictability of past events.”¹³² According to Martin Lijtmaer, hindsight bias stems from the psychological need to make sense of past events:

[W]hen people know an outcome, they naturally integrate the events leading to that outcome into a coherent story. In the process of constructing this narrative, people mentally emphasize certain circumstances that support the development of the known outcome while downplaying other circumstances that would have led to alternative plausible outcomes.¹³³

As it relates to Knobe’s theory, hindsight bias may cause people to overestimate the predictability of morally bad outcomes, thus leading them too often to conclude that they were subjectively intended.¹³⁴ An observer, believing that a person foresaw the

131. A fourth bias is the “Actor-Observer Bias,” which suggests that people “hold other[s] to different moral standards than [they] would hold [themselves] even if [they] were in the same situation.” See Thomas Nadelhoffer & Adam Feltz, *The Actor-Observer Bias and Moral Intuitions: Adding Fuel to Sinnott-Armstrong’s Fire*, 1 NEUROETHICS 133, 133 (2008). Because the Actor-Observer Bias is very general and broadly applicable, I do not discuss it in detail. However, as in any judgment about culpability or intent, the Actor-Observer Bias likely plays a role in felony murder cases.

132. Martin Lijtmaer, *The Felony Murder Rule in Illinois: The Injustice of the Proximate Cause Theory Explored Via Research in Cognitive Psychology*, 98 J. CRIM. L. & CRIMINOLOGY 621, 638 (2008) (citing Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 571 (1998)).

133. Lijtmaer, *supra* note 132, at 639 (citing Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 584 (1998)).

134. This problem also affects the primary utilitarian justification for felony murder (that it deters people from committing felonies that they know will endanger someone’s life) because observers likely overestimate the foreseeability of the danger. See, e.g., Baier, *supra* note 87, at 712–13 (citing 2 FRANCIS

consequences of her actions, might conclude that she subjectively intended them, but if the perception of foresight is incorrect, then the attribution of intent is incorrect as well.

Hindsight bias could thus undermine felony murder's retributivist justification. Lijtmaer explains that hindsight bias increases felony murder punishments by making juries "more likely to conclude that a death occurring during the course of a felony was foreseeable."¹³⁵ The Knobe Effect would exacerbate this problem by leading people to believe that the death was not only foreseeable but also intentional.

On the other hand, the problem of hindsight bias may simply suggest that states should use Binder's formulation of felony murder instead of Crump's. Recall that Crump's formulation restricts felony murder to crimes committed in ways that are "clearly dangerous to human life," whereas Binder's uses a predetermined list of "inherently dangerous" felonies.¹³⁶ I argued that Crump's formulation was better at ensuring that each defendant actually had sub-baseline intentions regarding the death,¹³⁷ but Crump's formulation would seem more prone to hindsight bias because it requires people to judge whether an action that caused a death was "clearly dangerous to human life." Because of the outcome, people will conclude too often that it was.

Binder's formulation may therefore better avoid hindsight bias in that restricting felony murder to a predetermined list of felonies removes the biased judgment from the equation. For example, no matter how strongly jurors believed that an online hacker intended to give her victim a heart attack, they could not convict her of felony murder if cybercrimes were not on the list of inherently dangerous felonies (as one would hope). Again, to offer maximum protection to defendants, lawmakers should probably adopt both Binder's and Crump's limitations, requiring that the underlying felony be included on the predetermined list *and* be perceived as having been committed in a way that was "clearly dangerous to human life." That combination would help to counteract hindsight bias and ensure that the rule punishes only intentional killings.

WHARTON, *WHARTON'S CRIMINAL LAW* 208 (1979); JOSHUA DRESSLER, *UNDERSTANDING CRIMINAL LAW* 464 (1987)). Hindsight bias, however, could suggest that people overestimate the foreseeability of the death, thus weakening the deterrence rationale.

135. *Id.* at 645. If the death was foreseeable, then the defendant is presumably more culpable, although Lijtmaer does not make this connection.

136. *See supra* Part II.C.

137. *See supra* Part II.C.

2. Motivational Bias

Similar to hindsight bias, motivational bias may suggest that the Knobe Effect represents a mistaken perception of mental states. Motivational bias is the phenomenon that people are *motivated to attribute* subjective intent to a person who caused a bad outcome.¹³⁸ When observers feel that the person who caused the outcome deserves blame for it, they are motivated to conclude that she subjectively intended it.¹³⁹ The Knobe Effect could thus be a result of an unjustified inclination to blame.¹⁴⁰

Knobe describes the psychological mechanism of motivational bias as an interaction between blame and the perception of subjective intent. Rather than leading directly from blame to intent, there is “a reciprocal relationship between people’s blame judgments and their intuitions about intention.”¹⁴¹ After initially feeling that a person is to blame, the observer reinterprets the situation to make the person

138. Knobe, *supra* note 5, at 321–24.

139. *Id.* This phenomenon has also been described in terms of the need to be seen as blaming the person for the morally bad outcome. See, e.g., Knobe & Burra, *supra* note 1, at 119–20 (citing F. Adams & A. Steadman, *Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding?*, 64 ANALYSIS 173–81 (2004); F. Adams & A. Steadman, *Intentional Action and Moral Consequences: Still Pragmatic*, 64 ANALYSIS 268–76 (2004)). In other words, the Knobe Effect might be explained by the desire to avoid being seen as letting a person off the hook for causing a morally bad outcome. *Id.* People would say that an action was intentional to make sure they are perceived as properly condemning the bad outcome. *Id.*

140. Knobe, *supra* note 5, at 321–24. Studies suggest that the Knobe Effect is not based on emotion, however. A study by Young showed that people who have lesions to the ventromedial prefrontal cortex (which means they have little or no emotional response) nonetheless exhibit the Knobe Effect. *Id.* at 322 (citing L. Young et al., *Does Emotion Mediate the Effect of an Action’s Moral Status on its Intentional Status? Neuropsychological Evidence*, 6 J. COGNITION & CULTURE 291–304 (2006)). In fact, people with these lesions exhibited the Knobe Effect more strongly, attributing intent to the CEO who causes harm to the environment 100% of the time (whereas only 87% of participants in the original study attributed intent in this scenario). *Id.* See also Knobe & Burra, *supra* note 1, at 117–18. Additionally, a study by Keys and Pizarro found that even when one actor is made to look like a “generally nice person” and another is made to look like a “generally nasty person,” the Knobe Effect continues to track the morality of the outcome, not the person. Knobe, *supra* note 5, at 322–23. An emotion-free Knobe Effect, however, would not necessarily exclude motivational bias because immediate moral appraisals can happen without any emotion at all. *Id.* at 322. But see Nadelhoffer, *supra* note 130, at 205–06 (citing “affective model[s] of moral psychology,” which hold that “emotional and non-rational processes, rather than deliberative and rational ones, are primarily responsible for our moral judgments”). Of course, the intersection of emotion, cognition, and moral judgments is a rich field, and space limitations preclude anything near a complete treatment.

141. Knobe, *supra* note 5, at 321.

appear more blameworthy and then feels even more strongly that the person is to blame. The observer then reinterprets the situation again to further justify the feeling. Any one of these reinterpretations, or a few of them in sequence, can make it seem as though the person subjectively intended the bad outcome.¹⁴² In this way, the initial desire to blame sets off a series of reinterpretations that lead observers to conclude, more often than they should, that bad outcomes were caused intentionally.¹⁴³

If the Knobe Effect is caused by motivational bias, then it does not suggest that felony murder can actually be restricted to intentional killings. The knee-jerk feeling that a person is to blame is not a rational way to evaluate intent. Retributive theory requires specific indicators of culpability to authorize punishment.¹⁴⁴

Knobe, however, rejects the argument that motivational bias underlies the Knobe Effect. In fact, he argues that the Knobe Effect has nothing to do with blame.¹⁴⁵ He cites a study involving two hypothetical situations: one in which morally bad conduct caused a morally good outcome, and one in which morally good conduct caused a morally good outcome. Because both outcomes were good, there was no need to blame. Nevertheless, people were more likely to call the action intentional when the conduct was bad.¹⁴⁶

This study suggests that judgments about *conduct*, more so than *outcomes*, affect the perception of intent. As Knobe says, “the data don’t actually suggest that people’s causal intuitions are being influenced by a judgment that the agent is to blame for the outcome. Instead, the data appear to suggest that these intuitions are being influenced by a judgment that the agent’s action itself is bad.”¹⁴⁷

142. *Id.*

143. As Nadelhoffer says, “[J]udgments concerning the moral blameworthiness or praiseworthiness of agents can have a similar influence on our ascriptions of intentional action.” Nadelhoffer, *supra* note 130, at 205 (citation omitted). The theory of motivational bias is also supported by a “growing body of evidence suggesting that people often adopt certain views as part of a post hoc attempt to justify prior moral intuitions.” Knobe, *supra* note 5, at 321 (citing P. Ditto et al., *Motivated Moral Reasoning*, in *THE PSYCHOLOGY OF LEARNING AND MOTIVATION* 307 (D.M. Bartels et al. eds., 2009); J. Haidt, *The Emotional Dog and Its Rational Tail: A Social Institutional Approach to Moral Judgment*, 108 *PSYCHOL. REV.* 814 (2001)).

144. *See supra* Part II.B.

145. Knobe, *supra* note 5, at 322 (citing C. Hitchcock & J. Knobe, *Cause and Norm*, 106 *J. PHIL.* 587 (2009)).

146. Knobe, *supra* note 5, at 323.

147. *Id.* (emphasis omitted). Indeed, if motivational bias were causing the Knobe Effect in this scenario, then people would have to be “displeased with the agent who performs the bad action, [have] their intuitions thereby become distorted by moral judgment, and . . . end up being motivated to conclude: ‘This bad guy must have been the sole cause of the wonderful outcome that resulted.’”

If the Knobe Effect is caused by judgments about conduct rather than outcomes, then it is not infected by motivational bias. Additionally, judging a person based on her conduct is consistent with Binder and Crump, who suggest that conduct, apart from mental states, is relevant to culpability.¹⁴⁸ In fact, the theories of Binder and Crump seem well suited to avoiding motivational bias in felony murder because their proposed limitations would force fact-finders to focus on the defendant's conduct and not the emotionally laden outcome.¹⁴⁹

3. Outcome Bias

Outcome bias, like the other two biases, may suggest that the Knobe Effect reflects a mistaken perception. Outcome bias is the phenomenon that judgments about a person's conduct are influenced by the outcomes it causes.¹⁵⁰ As explained in the previous section, motivational bias (dealing with blame for outcomes) is not a serious problem for Knobe's theory because research shows that people label actions intentional based not on outcomes but on the underlying conduct.¹⁵¹ Outcome bias, however, suggests that morally bad outcomes cause people to judge the *conduct* too harshly, thus making them attribute more intent to the action than

Id. If such an explanation is unlikely, then perhaps the Knobe Effect operates independently of judgments regarding blame. Moreover, some research suggests that people judge whether an action is intentional *before* they judge whether the actor is to blame for the outcome. *Id.* at 322. On the other hand, participants in that study may have assigned blame for the bad conduct regardless of the outcome.

148. See *supra* Part II.B. Binder, of course, argues that outcomes are relevant to culpability, Binder, *supra* note 59, at 1030–31, 1034, which is not necessarily inconsistent with avoiding motivational bias insofar as conduct and outcomes can be *prejudged* in the abstract. See *infra* Part III.C.

149. A problem with this argument is that felony murder cannot be divorced entirely from outcomes because it depends, by definition, on the fact that a death occurred. Moreover, as Binder argues, perhaps society should punish completed attempts more severely than failed attempts. Binder, *supra* note 59, at 1028. Nevertheless, considering that felony murder does not require proof of a subjective intent to kill or recklessly endanger, it is arguably less infected by motivational bias than the traditional murder rule. In other words, and as discussed in more detail in the next section, felony murder depends on the culpability of objectively observable action, not the subjective thoughts of the defendant, thus diminishing the effect of motivational bias.

150. Lijtmaer, *supra* note 132, at 641 (citing Philip G. Peters, Jr., *Hindsight Bias and Tort Liability: Avoiding Premature Conclusions*, 31 ARIZ. ST. L.J. 1277, 1282 (1999)) (explaining that outcome bias is “the tendency to judge the quality of a decision, good or bad, based on its consequences”).

151. Knobe, *supra* note 5, at 322–23. See also *supra* Part III.C.2.

they should. In this way, the morally bad outcome of death could cause observers to wrongly conclude that the underlying conduct embodied sub-baseline intentions.¹⁵²

Outcome bias thus presents a serious problem for my argument that felony murder can be restricted to intentional killings. All may agree that a killing is not intentional under Knobe's theory unless the conduct exhibited sub-baseline intentions regarding the death. I have argued that this is the proper limit for felony murder—in accordance with Solan, Binder, and Crump—because it ensures that the rule applies only to intentional killings.¹⁵³ However, people may be unable to judge whether a defendant's conduct exhibited sub-baseline intentions because the morally bad outcome biases their judgments about the conduct itself.

In a sense, however, one can *prejudge* a person's subjective intent regarding certain outcomes. For example, before the CEO implements her business plan, one can decide in the abstract whether her ambivalence toward the environment is above or below the baseline of expected intentions. In the same way, one can decide in the abstract which felonies, or which kinds of conduct during felonies, demonstrate sub-baseline intentions regarding the risk of death.

Perhaps, then, felony murder laws can avoid outcome bias by prejudging, in the abstract, which conduct demonstrates the kinds of sub-baseline intentions that make the homicide intentional. Even if outcome bias would cause a judge or jury to perceive the defendant's conduct as morally worse than it was, the law can remove such bias by limiting felony murder to conduct that satisfies a set of *predetermined* conditions indicating the requisite sub-

152. Worse still, outcome bias is strongest when the outcome is the worst, which suggests that people's judgments about conduct could hardly be more biased than in the case of an innocent person's death. As Lijtmaer explains, "research suggests that [outcome bias] is amplified as the severity of the injury increases." Lijtmaer, *supra* note 132, at 642 (citing Robert A. Caplan et al., *Effect of Outcome on Physician Judgments of Appropriateness of Care*, 265 J. AM. MED. ASS'N 1957, 1960 (1991); Philip G. Peters, Jr., *Hindsight Bias and Tort Liability: Avoiding Premature Conclusions*, 31 ARIZ. ST. L.J. 1277, 1283 (1999) ("Outcome bias appears to be most serious when the victim's injuries are severe. Although the research findings have been inconsistent, most conclude that severity is associated with a greater assessment of fault.")). Lijtmaer cites a study involving an actor who caused an injury. When the scenario was changed from a permanent injury to a temporary injury, participants judged the actor to have taken appropriate care 28% more often. Yet when the scenario was changed from a temporary injury to a permanent injury, participants judged the actor to have taken appropriate care 31% less often. *Id.* at 642 (citing Robert A. Caplan et al., *Effect of Outcome on Physician Judgments of Appropriateness of Care*, 265 J. AM. MED. ASS'N 1957, 1960 (1991)).

153. See *supra* Part II.C.

baseline intentions. Again, 35 states use lists of “inherently dangerous” felonies for this purpose, and Crump’s formulation—which looks for conduct during a felony that is “clearly dangerous to human life”—could serve the same purpose.¹⁵⁴ Outcome bias could thus be avoided by deciding in advance which types of conduct during a felony indicate the sub-baseline intentions that make the killing intentional.¹⁵⁵

Three problems remain. First, as Tomkovicz and Gerber point out, politics will influence which felonies and types of conduct are included on the list, likely making the list too long.¹⁵⁶ Second, outcome bias will still influence which felonies and types of conduct are included. The study cited by Lijtmaer, after all, involved hypothetical scenarios, and the participants still exhibited significant outcome bias.¹⁵⁷

Third, sentencing decisions are unavoidably infected with outcome bias. This is perhaps the most difficult problem to resolve. To be sure, outcome bias infects sentencing in all cases, not just felony murder, but the fact that felony murder authorizes murder-level punishment even when the defendant did not subjectively intend to kill or recklessly endanger exacerbates the problem of inflated sentencing. In other words, felony murder allows judges, who are unavoidably affected by bias, to be too harsh.¹⁵⁸ Moreover, although lawmakers can prejudice felonies and conduct, they cannot prejudice sentencing decisions because determinate sentences are

154. For example, lawmakers could prescribe certain kinds of conduct, similar to aggravating factors, required for a felony murder conviction.

155. Whatever role outcome bias plays in the Knobe Effect would be negated as well because, even if people wanted to attribute intent based on a biased perception of Solan’s baseline, they would not be able to impose punishment for intentional murder unless the action satisfied the predetermined conditions.

156. See *supra* Part III.A.

157. Lijtmaer, *supra* note 132, at 642. Perhaps this problem could be mitigated by structuring the deliberation process to avoid referring to the outcome of the conduct. Of course, the purpose of the discussion—to create rules for felony murder—would remain a biasing factor, but hopefully the specter of death would be sufficiently obscured to reduce the effect of outcome bias. Again, this discussion underscores how important it is for judges and juries *not* to have discretion to decide whether a certain instance of conduct constituted disdain for human life. When a real person’s death is before the decision-maker, outcome bias will be the strongest, which means that judges and juries will be most prone to overestimate the defendant’s culpability.

158. See, e.g., Douglas W. Schwartz, Note, *Imposing the Death Sentence for Felony Murder on a Non-Triggerman*, 37 STAN. L. REV. 857 (1985); Greg Bailey, Student Work, *Death by Automobile As First Degree Murder Utilizing the Felony Murder Rule*, 101 W. VA. L. REV. 235 (1998).

unconstitutional.¹⁵⁹ Outcome bias in sentencing, therefore, merits further consideration, especially with regard to felony murder.

V. CONCLUSION

I argue that Knobe's theory provides a framework in which lawmakers can restrict felony murder to intentional killings, thus confining the rule to the most deserving defendants and answering the retributivist charge that felony murder does not account for culpability. Yet the question remains: Should theories like Knobe's have any bearing on how criminal laws are designed, or should the law be structured on a purely rational basis, not bending to folk research about the popular understanding of words?

At one level, the question is beyond the scope of this Article, which simply presumes that retributive theory should at least consider what people actually mean when they talk about core concepts such as intent and culpability. Moreover, I avoid any discussion of utilitarian criminal theory, taking no position on the separate question of whether retributivism itself may be outdated, brutish, or ineffective at accomplishing the goals of criminal law. I strongly believe, however, that lawmakers and theorists should not ignore the current explosion of psychological and cognitive research that bears on concepts relevant to criminal law. Much of this research, like Knobe's, draws into question even the most established and widely accepted premises. Such is the case, moreover, under all theories of punishment: retributive, utilitarian, or otherwise. This Article is thus part of the growing movement that employs experimental research to examine and reevaluate the presumptions and justifications underlying legal doctrines. In the context of criminal laws, which routinely and as a matter of course deprive people of their most basic liberties, such continual reexamination could hardly be more important.

159. *United States v. Booker*, 543 U.S. 220, 249–258 (2005) (striking down the portion of the Federal Sentencing Act that created a determinate sentencing scheme).