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

Joseph Mauro
INTENTIONAL KILLING WITHOUT INTENDING TO KILL: 
KNOBE’S THEORY AS A RATIONAL LIMIT 
ON FELONY MURDER

Joseph C. Mauro*

I propose that Joshua Knobe’s theory of intentional action provides a rational way to limit felony murder so that it no longer offends traditional notions of retributive justice. Knobe’s research shows that when people view an action as morally bad, they are more likely to call it intentional. Based on this research, Knobe posits that intent refers not only to a mental state, but also to whether a person’s conduct, and the outcomes it causes, are morally bad.

A common retributivist criticism of felony murder is that it does not require proof of intent. Under Knobe’s theory, however, some instances of felony murder are actually intentional, even if the defendant did not subjectively intend to kill or recklessly endanger. That is, the combination of morally bad conduct (a dangerous felony), a morally bad outcome (death), and a morally bad mental state (willingness to risk death) makes the action as a whole intentional. If lawmakers were to restrict felony murder to such “intentional” killings, they would not only answer retributivist critics, but also eliminate the rule’s most troublesome applications.

* Law Clerk, Hon. Sally D. Adkins, Court of Appeals of Maryland.
INTRODUCTION.................................................................3

I. The Retributivist Criticism: Felony Murder Ignores Gradations of Culpability.................................................6

II. Argument: Felony Murder can be Limited to Intentional Killings, and thus Justified under Retributive Theory........9
   A. Knobe’s Theory: Intent Describes Morality.................10
   B. Retributive Theory: Intent Justifies Murder Punishments..........................................................17
   C. Limiting Felony Murder to Intentional Killings...........23

III. Counterarguments.........................................................28
   A. Felony Murder is Especially Corrupt and Unfair.......28
   B. Knobe’s Theory does not Apply to Unforeseen Outcomes..........................................................31
   C. The Knobe Effect Represents a Mistaken Perception..........................................................35
      1. Hindsight Bias.......................................................36
      2. Motivational Bias....................................................38
      3. Outcome Bias......................................................41

CONCLUSION........................................................................44
INTRODUCTION

Consider this scenario. The CEO of a major corporation has developed a new business plan. 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 into her office 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 business 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 into her office looking excited. “Not only is your plan going to make a lot of money,” he says, “it’s also going to end up saving 200,000 acres of rainforest that would have been destroyed otherwise.”

“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, one could say that 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 presented with 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

² Some researchers define the Knobe Effect more narrowly, applying it only to Joshua Knobe’s 2003 study. Research involving the phenomenon has expanded greatly since 2003, however. This article uses the term “Knobe Effect” to refer
According to some researchers, the Knobe Effect proves that “intent” refers not only to a person’s mental state, but also to her conduct and the outcomes it causes. On this view, when conduct, outcomes, and the person’s mental state are sufficiently bad, the action is intentional. Not all experts hold this view, however. Some argue that intent is nothing more than a mental state, and that the Knobe Effect represents a mistaken perception of a person’s mental state. Nevertheless, Joshua Knobe, the author of the study involving the CEOs, argues that the Knobe Effect has expanded the meaning of intent to include conduct and outcomes. An action is intentional, he claims, by virtue of morally bad conduct and outcomes, even if the person did not subjectively desire that such outcomes would occur. His is an objectively observable conception of intent.

Knobe’s theory provides a partial response to 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. Whereas traditional murder requires the prosecution to prove that the defendant had a certain mental state—whether “premeditation and deliberation,” “intent to kill,” or “recklessness”—felony murder requires none of these things. Instead, in cases of felony

to the broader phenomenon by which the morality of an action affects people’s perception of whether it was intentional. See Part II.A, infra.

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 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 Moralist, 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 that the person subjectively intend to bring it about or even risk it.

8 40 C.J.S. Homicide § 51 (2011)

murder, most jurisdictions presume that the defendant had the mental state required for traditional murder, simply because she committed a dangerous felony that caused a person to die.\textsuperscript{10}

Retributivist critics complain that felony murder does not tailor punishment to the defendant’s culpability. They maintain that intent determines culpability, and therefore that felony murder is illegitimate because it does not require proof of intent.\textsuperscript{11} Without proof of intent, the state cannot judge whether the defendant is culpable enough to merit murder punishments.\textsuperscript{12}

Knobe’s theory addresses this criticism by suggesting that lawmakers can amend the felony murder rule to require proof of intent, because certain cases of felony murder actually do involve intentional killings.\textsuperscript{13} Even if the defendant did not mean to kill, the killing was intentional if the defendant’s mental state and conduct were sufficiently immoral (taking for granted that the

\textsuperscript{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 fact that the felony is dangerous is important, because most states restrict felony murder to felonies that are clearly or inherently dangerous, with the idea that in such cases the death was foreseeable. See Birdsong, supra note 20, at 21, n. 140. This limitation is important for Knobe as well, because if the killing is purely accidental, people do not make the moral judgment that causes them to label it intentional. See Parts II.C., III.A., infra.

\textsuperscript{11} 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 C.J.S. Homicide §§ 73, 75 (2011).

\textsuperscript{12} Crump, supra note 11, at 1159–60. But see Part II.B, infra, discussing whether subjective intent is the only component of culpability.

\textsuperscript{13} I do not mean to suggest that concept of intent is uniform. To the contrary, legal doctrines vary with respect to how they define and use the concept of intent. See, e.g., McClesky v. Kemp, 481 U.S. 279, 293–95, n. 12 (1987) (discussing the difference between the intent requirements in Title VII cases, capital cases, and voting rights cases). My argument focuses on the linguistic meaning of intent, i.e., what people 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”).
outcome of death is maximally immoral). If states were to restrict felony murder to such cases, then it would no longer fall subject to the retributivist charge that it ignores intent, and would be better tailored to the most culpable defendants. Part I articulates the retributivist criticism of felony murder. Part II describes how Knobe’s theory addresses it. Part III deals with three counterarguments.

I. The Retributivist Criticism:
Felony Murder Ignores Gradations of Culpability

Retributivist scholars commonly criticize felony murder for departing from traditional notions of culpability. Retributive theory bases the level of a defendant’s punishment on her culpability, which is demonstrated by her intent. The more culpable the defendant, the more punishment is deserved. The retributivist argument against felony murder is that it ignores gradations of culpability because it does not require proof of the defendant’s intent. 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 it—but felony murder ignores intent entirely. Thus, many retributivist scholars consider the rule unjustified.

---

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


16 *See generally*, ROBINSON & DARLEY, *supra* note 15.


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 requirement for murder, whatever it may be in each jurisdiction, because the defendant committed a felony that caused a death. The felonious action that caused the death proves that that defendant had an “evil mind,” which is sufficient proof of her 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. In other words, the “constructive malice” rationale is
Indeed, felony murder is the only version of murder that does not require proof of any particular mental state.\textsuperscript{20} Traditionally, a homicide is not murder unless the prosecution proves that it was done with “premeditation and deliberation,” “recklessness,” “malice aforethought,” “indifference to human life,” or any of a number of highly culpable mental states.\textsuperscript{21} In cases of felony murder, on the other hand,

[T]he 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 defendant [knew] that his or her acts [would] cause death or great bodily harm to the victim.\textsuperscript{22}

\begin{footnotesize}
\begin{enumerate}
\item Indeed, felony murder is the only version of murder that does not require proof of any particular mental state.\textsuperscript{20}
\item Traditionally, a homicide is not murder unless the prosecution proves that it was done with “premeditation and deliberation,” “recklessness,” “malice aforethought,” “indifference to human life,” or any of a number of highly culpable mental states.\textsuperscript{21}
\item 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 defendant [knew] that his or her acts [would] cause death or great bodily harm to the victim.\textsuperscript{22}
\end{enumerate}
\end{footnotesize}
Additionally, the prosecution does not need to prove the underlying felony. Instead, 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.²⁴ The Model Penal Code, for example, appears to have abolished felony murder because it lacks an intent element.²⁵ Similarly, James Tomkovicz argues that felony murder is unjustified because it does not require proof of the defendant’s

²³ 40 CJS Homicide § 51. A few states, however, do not allow convictions on felony murder in the first degree without proof of malice or premeditation and deliberation. Id n. 29–32.
²⁴ Crump, supra note 11, at 1159–60, n. 10.
²⁵ Id. at 1159–60, 1164 (citing MODEL PENAL CODE § 210.2 cmt. (1982) As Crump explains, however, the MPC appears to have phrased 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 person 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 rule while actually preserving it in an altered form.’”); see also Part III.A, infra, discussing the similarity between the “abrogation” of the felony murder rule by the Michigan Supreme Court in People v. Aaron and the limitation of the felony murder rule to killings that are intentional under Knobe’s theory.
mental state.\textsuperscript{26} Retributive criminal theory, he says, depends upon “the idea that punishment should be commensurate with \textit{mental} fault.”\textsuperscript{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.\textsuperscript{28}

The court is likely correct that, “without more,” the fact that a killing occurred during a felony is not a sufficient indicator that the defendant deserves murder-level punishment. Traditional retributive theory requires proof of intent to show culpability. Thus, to bring felony murder in line with retributive theory, it would need to be limited to cases in which the killing was, in fact, intentional.

\textbf{II. Argument: Felony Murder can be Limited to Intentional Killings, and thus Justified under Retributive Theory}

According to Knobe, the Knobe Effect suggests that intent refers not only to a mental state, but also to the morality of conduct

\textsuperscript{26} Tomkovicz, \textit{supra} note 11, at 1441.
\textsuperscript{27} Id (emphasis added).
\textsuperscript{28} People \textit{v.} Aaron, 409 Mich. 672, 703 (1980); \textit{see also} Crump, \textit{supra} note 11, at n. 10 (listing, among those who advance this criticism of felony murder, “Robert M. Elliot, \textit{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) (arguing that ‘the rule does violence to the philosophy which dictates that criminal liability should be commensurate with moral culpability’); George P. Fletcher, \textit{Reflections on Felony Murder}, 12 SW. U. L. REV. 413, 427–28 (1981); Jeanne H. Seibold, \textit{The Felony Murder Rule: In Search of a Viable Doctrine}, 23 CATH. L\textsc{aw} 133, 160–61 (1978) (asserting that the rule is ‘grossly misplaced in a legal system which recognizes the degree of mental culpability as the appropriate standard for fixing criminal liability’ and that abolishing the rule would lead to blameworthiness as the guide for imposing punishment’)).
and outcomes.\textsuperscript{29} If this is true, then some cases of felony murder involve intentional killings, because the immoral conduct (a dangerous felony) and outcome (a death), combined with a sufficiently immoral mental state regarding the death (a willingness to risk it), makes the action as a whole intentional.\textsuperscript{30} In fact, the indicators of intent under Knobe’s theory (conduct, outcomes, and mental state) are the same factors that some retributivist scholars have used to suggest ways in which felony murder can be limited to the most culpable defendants.\textsuperscript{31} Thus, I propose that properly limited versions of the felony murder rule, far from undermining the retributive connection between intent and culpability, actually support it.

A. Knobe’s Theory: Intent Describes Morality

Following Knobe’s groundbreaking study, subsequent research confirmed and expanded his findings.\textsuperscript{32} For example, studies found that the Knobe Effect occurs in scenarios that use terms such as “intention,” “deciding,” “desire,” “in favor of,” “in order to,” and “advocating” to refer to intent.\textsuperscript{33} It also occurs at what might be called the edges of intent; for example, in the scenario involving the CEOs, 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.”\textsuperscript{34} Another study showed that the Knobe Effect occurs even when the person did not know that the bad outcome would result, but

\textsuperscript{29} Id.
\textsuperscript{30} See Part III.C, \textit{infra}, discussing which mental states are required to call a killing intentional under Knobe’s theory.
\textsuperscript{31} See Part II.B, \textit{infra}.
\textsuperscript{32} Recall that the Knobe’s study about the CEOs revealed that people are more likely to say a person intentionally caused a morally bad outcome when their conduct was morally bad as well. Knobe & Burra, \textit{supra} note 1, at 118–19. When the CEO harmed the environment, 87\% of people said that she did so intentionally, but when the CEO helped the environment, only 20\% of people said that she did so intentionally. Id.
\textsuperscript{33} Knobe, \textit{supra} note 5, at 318; \textsc{Joshua Knobe & Shaun Nichols}, \textsc{Experimental Philosophy} 137 (2008). For example, 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.’” Id. at 319.
\textsuperscript{34} Id.
nonetheless wanted it to happen.\textsuperscript{35} Thus, for purposes of this article, “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 \textit{subjectively intending} something is distinct from the broader concept of “acting intentionally,”\textsuperscript{36} which includes the morality of conduct and outcomes in addition to mental states.\textsuperscript{37} In other words, although subjectively intending is one kind of intent, it is not the only kind; immoral conduct that causes bad outcomes is another kind of intent.\textsuperscript{38}

If Knobe is correct that morally bad conduct and outcomes make an action intentional, then felony murders are often intentional.\textsuperscript{39} When a dangerous felony causes a death, the moral

\textsuperscript{35} Nadelhoffer, \textit{supra} note 4, at 345–46. See also Part III.B, \textit{infra}, discussing the possibility that the Knobe Effect does not occur unless the person knew that his actions would cause the relevant outcome.

\textsuperscript{36} Knobe & Burra, \textit{supra} note 2, at 124.

\textsuperscript{37} \textit{But see} Knobe, \textit{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 Part III.C, \textit{infra}, discussing whether the perception of intent in the Knobe Effect adds to the meaning of intent (Knobe’s theory) or is simply a mistaken perception of the person’s mental state.

\textsuperscript{38} As Knobe says, there is no logical reason to assume that “a theory of intentional action needs to be related to a conception of future intention.” Id at 2.

\textsuperscript{39} Id. Again, this assumes that the defendant had some culpable mental state, short of intent to kill or recklessly endanger, regarding the death. \textit{See} Part II.C, \textit{infra}. \textit{But see} Lawrence M. Solan, \textit{Blame, Praise, and the Structure of Legal Rules}, 75 BROOK. L. REV. 526–27, 529 (2009), (discussing Steve Guglielmo & Bertram Malle’s explanation of the Knobe Effect, and referring to a study by Steven Sverdlik, which discovered that, “when given the opportunity, people are perfectly willing to find conduct both blameworthy and unintentional at the same time”); Knobe, \textit{supra} note 5, at 339 (referencing Frank Hindriks’ theory that what motivates the Knobe Effect is the person’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”).
badness of the conduct and the death can make the entire action intentional, even if the person did not subjectively intend to kill.40

The connection between Knobe’s theory and felony murder becomes clearer in light of Lawrence Solan’s description of the mechanics of the Knobe Effect. Solan hypothesizes that there is a baseline of subjective intentions that people expect of each other. 41

In general, people assume that other people intend good outcomes.42 When a person fails to intend good outcomes—for example, when the CEO does not care about affecting the environment—that person has fallen below the baseline of expected intentions.

Solan argues that only morally bad actions need to be described as intentional, because people simply assume, based on the baseline, that good actions were subjectively intended. 43 He refers to Paul Grice’s “Maxim of Quantity,” which holds that

---

40 On the other hand, some research suggests that morality and culpability are divorced from the perception of intent. See, e.g., Knobe, supra note 5, at 341–42 (wherein Tamar Kreps and Benoit Monin argue that while the baseline theory may be true, the baseline depends not only on morality, but also on “prescriptive norms (how we think the group believes people should act) and descriptive norms (How we think group members actually act, regardless of how they should),” as well as moral judgments); id. 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 actually explains why the Knobe Effect occurs, since 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 2, 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) (citing Bratman 1984, 1987; Mele 1989, 1992); accord Solan, supra note 42, 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 that the “CEO intentionally harmed” the environment, but 70% said that the “CEO intentionally adopted a profit-raising program that he knew would harm the environment,” and that the respondents assigned blame to both groups).

41 There may be a normative element to these expectations, but as described in Solan’s piece, they are primarily predictive. See Solan, supra note 42, at 522–28

42 Id.

43 Id at 522.
dialogue should only include relevant statements.\textsuperscript{44} Thus, he says that people need not mention the intent behind an action unless it contradicts the presumption that people intend good outcomes.\textsuperscript{45} For purposes of this article, Solan’s theory of the mechanics of the Knobe Effect’s will be called the “baseline theory.”

Solan proves his baseline theory with a thought experiment involving Mario and a printer.\textsuperscript{46} When someone says, “Mario fixed the printer,” there is no need to say that he did so intentionally. People presume that he intended to perform the action because it was good. Similarly, 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 sub-baseline intentions cause morally bad outcomes—for example, when they learn that the CEO did not care about the environment and then harmed it. Unlike Mario, the CEO fell below the baseline of expected intentions because she did not have good intentions regarding the environment. Thus, when the environment was harmed, people had a reason to say something about her intent. People noted her sub-baseline intentions, and expressed disapproval of them,\textsuperscript{47} by saying that she caused the harm intentionally.\textsuperscript{48} Importantly, however, sub-baseline intentions do

\textsuperscript{44} Id.
\textsuperscript{45} Id.
\textsuperscript{46} Id.
\textsuperscript{47} See Part II.C, infra, discussing the hypothesis that the state must condemn antisocial values.
\textsuperscript{48} Solan, supra note 42, at 524. It is also important to remember here that the question of intent is not binary, but rather operates 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 represent the agent as strongly opposed, as strongly in favor, or as occupying any of the various positions in between.

Such a notion comports with the Model Penal Code’s gradations of intent. See Model Penal Code §§ 210.0-210.4 (1985). On the other hand, Joseph Ulatowski and Justus Johnson, in response to Knobe, argue that while it might make sense to talk about a baseline of expected intentions, the theory is not valid unless the baseline is more clearly defined. Knobe, supra note 5, at 352–53
not indicate intentional action unless they have to do with the morally bad outcome. Sub-baseline intentions about other things—for example, if the CEO intended to steal her neighbor’s car after work—are not relevant to the question of whether she intentionally harmed the environment.\footnote{See, Knobe, supra note 5, at 322–23 (citing a study by Keys & 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).}

Solan’s baseline theory finds support in Knobe’s theory of alternative possibilities. Knobe’s alternative possibilities are the morally good actions that people think others should perform and the morally good intentions they think others should have.\footnote{Knobe, supra note 5, at 326} When a person causes a bad outcome, people judge her in relation to their expected good actions and intentions.\footnote{Id.} In other words, moral judgments about what should have happened influence how people “make sense” of what actually happened—including whether an action was intentional.\footnote{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 is beginning to suggest that moral intuitions influence how people perceive causation and permission, in addition to how they perceive intent. See, e.g., Knobe, supra note 5, at 319 (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 42, at 530–31 (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 skewing of people’s perception of causation would seem to exacerbate the perceived injustice of felony murder by punishing not only defendants who lack the subjective intent to kill, but also those who did not even cause, at least in the traditional sense, the}
Solan’s baseline theory and Knobe’s alternative possibilities theory help explain why certain felony murders are intentional. Certain conduct contradicts the presumption that people intend to avoid killing other people. For example, a person falls below the baseline of expected intentions 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 their expected alternative possibilities, sharpening their moral condemnation. In this way, 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 intent. As with the CEO, they say the killing was intentional because the conduct, outcome, and mental state were sufficiently bad.  

Knobe and Solan were not the first to put forward a theory of intent that helped to justify 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 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).  

One may ask at this point what distinguishes my version of felony murder from depraved heart murder, which punishes people for accidentally killing 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, because it simply describes depraved heart murder in which the killing occurs during a dangerous felony. The addition of the felony is important, however, because it ensures that the conduct, one of Knobe’s three keys to intent, is 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 severity of the badness with the tendency to call the action intentional.
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.’

As this passage demonstrates, even 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. Although Knobe employs new empirical research to support the claim that intent describes morality, the idea is not new.

In fact, Knobe’s theory parallels the centuries-old concept of “malice aforethought,” one of the traditional elements of murder. In the Thirteenth Century, “malice aforethought” did not refer to a subjective intent to kill or recklessly endanger, but rather “appears to have meant nothing more definite than a general intention to commit a wrongful act.” 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.” Thus, it appears that the law of homicide has always made room for the notion that a person is guilty of an intentional killing when, instead of subjectively intending to kill, a dangerous felony causes a death.

---

54 Tomkovicz, supra note 11, at 1471–72. Of course, the fact that something is not an accident does not necessarily mean that it was intentional. However, something that is accidental is never intentional. Thus, the refusal of lawmakers view felonious deaths as accidental makes it possible to call them intentional.

55 Id.

56 Comment, Felony Murder as a First Degree Offense: An Anachronism Retained, 66 YALE L.J. 427, 430 (1957) (citing 2 Pollock & Maitland 480–81, 469; Rollin M. Perkins, A Re-Examination of Malice Aforethought, 43 Yale L.J. 537, 545 (1934)).

57 Id. (citing Note, 59 Dick L. Rev. 183, 185 (1955); Note, 31 Ind. L.J. 534, n.3 (1956) (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 her 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 felon as if she had the subjective intent traditionally required for murder, even though likely did not have such subjective intentions. See note 19, supra, for further discussion.
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 Thirteenth Century sensibilities, then perhaps it is not relevant to contemporary murder doctrines. Modern criminal law is explicitly based upon gradations of *subjective* intent.\(^{58}\) By moving away from subjective intent, a conception of felony murder that relies on the Knobe Effect could appear outdated or backwards.\(^{59}\) Yet some argue that modern criminal law’s reliance on subjective intent fails to adequately address the culpability of felony murders.\(^{60}\) Showing that some felony murders are intentional, then, could reconcile the two positions, as both sides agree that intent demonstrates culpability and culpability justifies punishment.

**B. Retributive Theory: Intent Justifies Murder Punishments**

When a killing is intentional, retributivists should support imposing murder-level punishments, because the retributive theory bases punishments on culpability and culpability on intent.\(^{61}\) Some


\(^{59}\) 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 63, at 430–32. This notion was first expressed by the Pennsylvania legislature following the Revolutionary War as “willfulness, premeditation and deliberation . . . constituting in combination a design to kill directed against a selected individual.” Id. Second degree murder, on the other hand, may be reckless or negligent, which means that felony murder must necessarily be in the second degree because the defendant did not subjectively intend to kill. Id. at 431–32. This is really an argument for the abolition of first degree felony murder. For purposes of this article, however, it at least shows that the connection between historical notions of malice and the Knobe Effect may not explain the modern felony murder rule as much as it might otherwise appear. Of course, felony murder laws also find justification in utilitarian theories and other theories of punishment. They are not based solely on the theory of culpability embodied in Solan’s baseline theory. *See also* Part III.C, *infra*, discussing the possibility that the Knobe Effect is an undesirable part of our psychology that the law should push us to abandon in favor of more progressive and rational ways of thinking.

\(^{60}\) [cite crump and binder, generally]

retributivist scholars have argued that intent, for purposes of determining culpability, refers only to the defendant’s mental state. As David Crump observes, these scholars believe that “mens rea is the only legitimate determinant of blameworthiness.” Indeed, even when murder statutes do not require a subjective intent to kill—for example, statutes based on “gross recklessness . . . of a sufficient magnitude to evince a callous or depraved indifference to the value of human life”—they are still based solely on the person’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 theory has truly deepened the meaning of intent, then a person’s culpability should be judged by all aspects of her intent, including conduct and outcomes.

See, e.g., Tomkovicz, supra note 11, at 139–40 (arguing that the law should impose a “gradation [of punishments] proportionate to the established level of mental fault”); Binder, supra note 66, 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)).

Crump, supra note 11, at 1159.

Tomkovicz, supra note 11, at 139–40.

See Part III.C, infra, discussing whether this is the correct interpretation, or whether the Knobe Effect simply represents a mistaken perception of the person’s mental state, still the only component of intent.

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 judges and juries more likely to say that a defendant subjectively intended to kill or recklessly endanger, even when the defendant may not have so intended. See Part III.C, infra. Therefore, even though the perceived culpability of the defendant in felony murder cases may be inaccurate, such inaccuracies infect traditional murder convictions as well. Indeed, perhaps they infect traditional murder convictions more than felony murder cases, because in traditional murder cases the only way to determine intent is through the impossible task of determining what was going on in the defendant’s mind.
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 uses to describe intent. Their position is that conduct and outcomes demonstrate a person’s culpability, which is consistent with Knobe’s theory that conduct and outcomes demonstrate intent. Importantly, by describing the connection between culpability, conduct, and outcomes, Crump and Binder unknowingly sketched a framework for restricting felony murder to killings that would be labeled intentional under Knobe’s theory.

Crump clearly 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.

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” her values. Criminal punishments, therefore, must address the “expression” and “gratification” of bad values in the same way that they address culpable mental states. Binder calls this the “expressive theory” of culpability.

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

67 See Crump, supra note 11, at 1159–60, 1162; Binder, supra note 66, at 974–75.
68 See Crump, supra note 11, at 1159–60, 1162; Binder, supra note 66, at 974–75.
69 Knobe & Burra, supra note 1, at 123–125.
70 Crump, supra note 11, at 1159–60, 1162.
71 Binder, supra note 61, at 1030–31, 1034.
72 Id. at 974–75.
73 Id. at 970.
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.\textsuperscript{74}

Binder argues that values are important because the state’s social contract with its citizens requires punishing the expression of bad values. His idea is based on John Rawls’ theory that the purpose of the state is to “secure to each individual the broadest sphere of freedom compatible with like freedom for others.”\textsuperscript{75} For the state to operate well, it must ensure that citizens embody the kinds of values that enhance freedom for everyone. Therefore, he says, the state 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 has a duty to condemn those values through punishment.\textsuperscript{76}

\textsuperscript{74} Id. David Crump agrees with Binder, maintaining that blameworthiness depends on more than subjective intent. Crump, \textit{supra} note 11, 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.”).

\textsuperscript{75} Binder, \textit{supra} note 61, at 1018–19.

\textsuperscript{76} Id. at 974–75, 1025, 1032. Binder argues that the Model Penal Code’s causation standard reflects the intuition that bad values should be punished. Id. at 1025 (“The Model Penal Code causation standard restates the scope of the
Crucially for felony murder, Binder posits that even unexpected outcomes affect how much punishment is deserved. This is because outcomes, even if unexpected, are inextricably tied up with a person’s values; the worse the outcome, the stronger the expression of bad values, and thus the more punishment is deserved. In this way, the morality of a defendant’s conduct and the morality of the outcomes that she causes affect her culpability. Thus, if a defendant’s bad conduct causes a death, greater punishment is justified than when it does not, even if it is the same conduct with the same mental state.

It asks whether ‘the actual result differs from the probable [or designed] result only in the respect that a 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 . . . .” (citing JOSEPH RAZ, THE MORALITY OF FREEDOM 288–320 (1986); ELIZABETH ANDERSON, VALUE IN ETHICS AND ECONOMICS 6–7, 11–15 (1993)).

For example, Binder argues that a successful murderer deserves more punishment than someone who failed an attempted murder, even though their actions expressed the same bad values. See, e.g., Tomkovicz, supra note 11, 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 it 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
Binder and Crump agree, therefore, that a properly limited version of felony murder would be justified under the retributive theory.\textsuperscript{79} 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, a traditional murder.\textsuperscript{80} Similarly, Crump argues that felony murder is a necessary part of the criminal law because traditional murder, which cannot impose murder punishments without proof of a subjective intent to kill, does not account for the full range of factors that determine culpability, including conduct and outcomes.\textsuperscript{81} Thus, he argues that felony murder “may actually serve the policy of linking the criminal law to moral blameworthiness.”\textsuperscript{82}

Following Crump and Binder, I propose that the best way to ensure that felony murder punishes only those who are as culpable as traditional murderers is to restrict it to cases in which

\textsuperscript{79}Crump, \textit{supra} note 11, at 1159–60, 1162; Binder, \textit{supra} note 61, at 1030–31, 1034.
\textsuperscript{80} See generally Binder, \textit{supra} note 61.
\textsuperscript{81} Id.
\textsuperscript{82} Id. Moreover, perhaps evaluations of culpability should 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, \textit{e.g.} Christopher Slobogin, \textit{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.”). See also Part III.C, discussing the possibility that the Knobe Effect does not affect the meaning of intent, but rather is merely evidence of people’s inability to perceive others’ subjective intentions.
the killing would be called intentional under Knobe’s theory. In this way, felony murder would fill a gap in homicide schemes, addressing the set of intentional killings not covered by traditional murder while at the same time eliminating instances in which less culpable defendants are punished too harshly. If such a rule is desirable, the next question is how to structure it so that it applies only to intentional killings.

C. Limiting Felony Murder to Intentional Killings

Although Knobe’s theory suggests that felony murder sometimes involves intentional killings, certainly not all deaths that occur in the course of a felony are intentional. As 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, cannot affect the intentionality of the killing. Thus, purely unforeseeable or accidental killings should not be considered felony murders, 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. Under his “expressive theory,” a felony murder conviction is not justified unless the defendant had

---

83 See discussion supra note 101; see also Knobe, supra note 5, at 322 (describing a study suggesting that the perception of intent is tied to the morality of the conduct, not blame for the outcome that eventually results (citing Hitchcock & Knobe 2009)).

84 Binder, supra note 61, 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.
bad values that were “indulged” or “gratified” by the death.\textsuperscript{85} Thus, he argues that only inherently dangerous felonies, 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.\textsuperscript{86}

Most states appear to agree with Binder on this point. As of 2005, thirty five states had restricted felony murder to a list of felonies that pose the greatest risk to life.\textsuperscript{87} When a death results from one of these felonies, it is easier to say that the defendant expressed disdain for human life.\textsuperscript{88} Furthermore, as Crump

\textsuperscript{85} Id. at 1027–29, 1039–40.

\textsuperscript{86} Id. at 1028, 1058. An obvious counterargument is that even if a defendant wanted to expose someone to a risk of death in order to perpetrate a rape or robbery, the lack of subjective intent to kill suggests that the death may not gratify the defendant’s values, at least not to the extent that Binder argues.

\textsuperscript{87} Birdsong, supra note 20, at 21, n. 140; see also Tomkovicz, supra note 11, at 1467; Crump, supra note 8. Leonard 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 52 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 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”); see also Crump, supra note 11, 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.”). Though the sanctity of human life may not be one of the values necessary to enforce the Rawlsian social contract, it is nevertheless an important value for the law to uphold. On the other hand, the maximization of freedom and social good for all citizens does arguably depend on respecting everyone’s right to remain alive.

\textsuperscript{88} The development of common law felony murder in the United States also supports limiting the modern rule to inherently dangerous felonies, because the earliest forms of felony murder did not encompass accidental killings. Binder, supra note 22, at 203 (“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, \textit{Criminal Law-A Reckless Indifference to Human Life is Sufficient Evidence to
observes, the fact that a majority of states limits felony murder to inherently dangerous felonies shows that the version of the rule most commonly criticized by scholars—the version that punishes “any felony that results however unpredictably in a death”—is not representative of the general rule. 89

Crump, like Binder, would limit felony murder to crimes committed in a way that is “clearly dangerous to human life.” 90 This formulation attempts to restrict murder punishments to cases in which the defendant’s conduct expressed a high level of disdain for human life. 91 The difference is that Crump does not use a

89 Crump, supra note 11, at 1176–77. This is not to suggest, however, that individual judges and prosecutors cannot ignore such limiting language.
90 Id. at 1171 (criticizing California’s felony murder law, which is limited to inherently dangerous felonies in the abstract, because “[t]he 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[,]’”).
91 Id. at 1166 (arguing that limiting felony murder to felonies committed in ways that are “clearly dangerous to human life . . . ties the crime of murder to 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 (1984) (arguing that the felony murder rule would be unnecessary if all homicides were punished in proportion to the degree to which the felon exhibited disdain for the value of human life; “If the totality of the circumstances surrounding a homicide indicates that the defendant consciously subjected a human life to unreasonable risk, then murder punishments are justified. If the defendant’s conduct does not suggest a
predetermined list of “inherently dangerous” felonies, because even those felonies, in some cases, are committed in ways not clearly dangerous to human life. Thus, Crump’s formulation attempts to limit felony murder to cases in which each defendant actually expressed disdain for human life.

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 fall below Solan’s baseline regarding the possibility of death. 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 hold that the killing was intentional.

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

recklessness or an intent to kill, then the defendant’s culpability is insufficient to invoke murder penalties”).

92 Id.
93 Id.
94 But see Part II.B, supra, discussing whether culpability has traditionally been gauged by intent as defined by Knobe, or as defined simply as a mental states. It is impossible to answer this question in hindsight, because before Knobe’s study of 2003, there was no reason to think that the two concepts might be distinct.
95 Id. at 1166, 1175
96 Id.
Perhaps states should 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 felony is on the list of inherently dangerous felonies and was committed in a way that was 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

D. Cent. Code, § 12.1-16-01 (2005); Ohio Rev. Code Ann. § 2903.01 (2005); 21 Okla. Stat. 21, § 701.7 (2004); Or. Rev. Stat. § 163.115 (2003); R.I. Gen. Laws § 11-23-1 (2005); S.C. Code Ann. § 16-3-20 (2004); S.D. Codified Laws § 22-16-4 (2005); Tenn. Code. Ann. § 39-13-202 (2005); Tex. Penal Code § 19.03 (2005); Utah Code. Ann. § 76-5-201 (2005); Va. Code Ann. § 18.2-31 (2005); Wash. Rev. Code § 9A. 32.030 (2005); W. Va. Code § 61-2-1 (2005); Wyo. Stat. Ann. § 6-2-101 (2005)”). These states restrict felony murder to felonies that appear on a predetermined list of crimes typically dangerous to human life. See id. Of course, one can question whether the felonies 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 52, at 201–205. That is, the extent to which a certain value is “antisocial” and deserving of punishment depends on the particular values, Rawlsian or otherwise, of that jurisdiction. Id. It is interesting to consider, in this context, how felony murder might vary across jurisdictions with different values. Could a felony murder 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. Joshua Knobe has begun such research, but it is in its infancy. One study shows that the Knobe Effect operates in the Hindi language, but not necessarily in India itself. Knobe & Burra, supra note 1, at 126–29 (the subjects of the study were in the United States). Forthcoming research also suggests that the Knobe Effect is weaker among Asian American males. See Lawrence Ngo & Walter Sinnott Armstrong (forthcoming 2011). Moreover, it would appear difficult to discern whether the relevant subset is geographic or demographic, as the two categories cannot be entirely separated from each other. Binder would choose geography, arguing that the relevant values are those expressed by the law of each jurisdiction. Binder, supra note 61, 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, paper for American Society of Criminology Annual Meeting 30 October 1975.
could single out defendants culpable enough to deserve murder level punishment, and thus should find support in retributive theory.  

III. 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 number of intentional felony murders. Finally, some researchers argue that the Knobe Effect simply has no bearing on the meaning of intent, but instead represents a mistaken perception of a person’s mental state. If this is true, then felony murders are never intentional, and the Knobe’s theory cannot answer the retributivist criticism at all.

A. Felony Murder is Especially Corrupt and Unfair

If the felony murder rule operates in especially unfair or corrupt ways, then it may not be justifiable, even if it is 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. 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

---

98 This limitation also appears to track the traditional definition of “malice aforethought.” See Tomkovicz, supra note 11, at 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)).

hatred for crime onto a few unfortunate individuals. The nature of the political process exacerbates the problem, he says, because the fact that the defendant is a felon encourages lawmakers and judges to increase punishment to an unreasonable degree:

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.

Tomkovicz levies a similar charge, arguing that felony murder’s unfairness stems from the political process. Legislatures and judges, he argues, feel pressure to placate a populace that does not care about treating felons fairly. This pressure often overrides their ability to maintain “consistency and fairness” when charging and punishing felony murder defendants.

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

100 Id. at 782–84
101 Id. at 783 (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 will call a killing intentional if it involves conduct that exhibits sub-baseline intentions regarding the possibility of death. Gerber’s argument, however, is that people who fall below the baseline do not receive fair treatment. Id. at 782–84. In other words, it is not fair to apply inconsistent rules to the people below the baseline. To do so unfairly punishes felony defendants, “the very ones who need to learn something about principle.” Id. at 784.
102 Tomkovicz, supra note 11, at 1461–63, 1475.
103 Id.
104 Id.
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.  

Knobe’s theory could suggest, however, that these criticisms are misdirected. Gerber and Tomkovicz rail against applying felony murder to accidental killings, but as explained above, Knobe’s theory does not apply in such cases. Knobe’s theory requires the defendant to demonstrate sub-baseline intentions regarding the victim’s life before the killing can be called intentional. Thus, purely accidental killings cannot be called intentional, and cannot merit murder-level punishment.

Moreover, the principal case cited by Judge Gerber, People v. Aaron, actually supports the limited versions of felony murder that Crump and Binder suggest. While it is true that Aaron abrogated felony murder in Michigan, it expressly sanctioned murder-level punishments for defendants who caused deaths, without meaning to, while committing felonies that exhibited disdain for human life. The court observed that in such cases, felony murder is irrelevant because the prosecution can prove the intent required for murder.

Knobe’s theory, in accordance with thirty five states, says essentially the same thing: that 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

---

105 Id.
106 See id. at 763, 765, 767, 771, 778, 782–83; Tomkovicz, supra note 11, at 1438, 1444, 1449, 1472–73
107 See Part II.C, supra.
109 People v. Aaron, 409 Mich. 672, 730–31 (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 11, at 1159–60, 1164. Perhaps, then, the felony murder rule is superfluous in jurisdictions where the decision to commit an inherently dangerous satisfies the mens rea element of murder.
suggested by Crump and Binder). Thus, it is mostly a question of semantics whether Aaron “abrogated” felony murder or simply changed it to the version suggested by Crump and Binder. \(^{111}\) In this way, Knobe’s theory suggests that felony murder is not necessarily “unfair” or “capricious,” but is consistent with the meaning of intent when properly restricted.

B. Knobe’s Theory does not Apply to Unforeseen Outcomes

Some researchers maintain that the Knobe Effect occurs only when the person knows that she will cause the relevant outcome, \(^{112}\) which could suggest that Knobe’s theory does not apply to most felony murders because, in most cases, the killer does not know that her conduct will cause a death. \(^{113}\) Al Mele and Paul Moser, for example, argue that intent has nothing to do with situations in which the person does not believe that she is in control of the relevant outcome. \(^{114}\) 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. \(^{115}\)

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

\(^{111}\) Id; Crump, supra note 11, at 1159–60, 1164.
\(^{112}\) See generally Knobe, supra note 12 (summarizing various studies and disagreements about the Knobe Effect, all of which deal with scenarios in which the actor foresees morally good and bad outcomes).
\(^{113}\) 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 a relatively small percentage.
\(^{114}\) Nadelhoffer, supra note 4, at 341, 344.
\(^{115}\) Id.
morally bad outcome. The Knobe Effect, he says, is simply a condemnation of the person’s failure to avoid foreseen consequences:

[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.

If these researchers are correct, then perhaps Knobe’s theory is irrelevant to most felony murder cases. As with Lisa and Mike, most felony murderers do not believe that their conduct will cause a death. Instead, the death involves a significant amount of bad luck. A recent study by Nadelhoffer, however, suggests that the Knobe Effect actually does occur even when the person does not know or believe that her conduct will cause the relevant outcome. If this is true—if the Knobe Effect extends this far—then perhaps Knobe’s theory does speak to felony murder.

---

116 Knobe, supra note 5, at 339.
117 Id.
118 Id. Of course, I am proposing that Knobe’s theory might justify punishment for intentional killing even when the person did not have specific intent. Also, Hindriks’ statement reveals that Knobe disagrees with him, believing instead that an action can be intentional even when the person does not have foresight of the consequences. Id. See also Part III.B, infra, discussing “hindsight bias.”
119 Nadelhoffer, supra note 4, at 341. 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 . . . intentional action appear to be mistaken.” Id.
120 Crump also rejects the idea that felony murder should be restricted to cases in which the defendant subjectively believed that his action was likely to cause a death. Crump, supra note 11, at 1168. Crump argues that other kinds of murder do not require such precise subjective intent—for example, “depraved heart” murder. Id. at 1168.
Nadelhoffer’s study asked participants to examine four hypothetical scenarios and evaluate whether the person in each acted intentionally.\textsuperscript{121} 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.\textsuperscript{122} 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 grandmother to hasten his inheritance. To base the outcome on skill, the hypothetical stipulated that he was an expert marksman. To base it on luck, it stipulated that he had never fired a gun before.\textsuperscript{123}

The remarkable results involve Jake’s lucky shots. Even though he did not foresee the consequences of his conduct—he did not believe that he could hit his grandmother or the bull’s-eye—participants nonetheless called his action intentional more often when the outcome was morally bad.\textsuperscript{124} Seventy-six percent of

\begin{flushright}
\textsuperscript{121} Id. This study built upon a study by Knobe, which was the first to suggest that the Knobe Effect may not be limited to scenarios in which the actor foresees the consequences of her actions. Id.
\textsuperscript{122} Id.
\textsuperscript{123} Id. Nadelhoffer describes the scenarios in this way: “(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. 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 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. 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 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.
\textsuperscript{124} Id. at 345–46.
\end{flushright}
participants said that he intentionally shot his grandmother, while only twenty eight percent said that he intentionally shot the bull’s-eye. These results suggest that the Knobe Effect occurs even when the person does not know that her conduct will cause the relevant outcome.

On the other hand, Jake hoped that he would hit the targets, and he knew that doing so was possible. 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. Thus, while Nadelhoffer’s study may show that the Knobe Effect applies 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 possible, or in which the person does not desire the outcome. If the Knobe Effect does not apply in these situations, then it still may not apply in the majority of felony murder cases, considering that most defendants probably do not hope to kill and may not know that death is a possibility. As of this writing, no studies have tested whether the Knobe Effect applies in such situations.

125 Id. Unsurprisingly, the skillful scenarios also exhibited the Knobe Effect. When Jake was an expert marksman, 95% of respondents said that he shot his grandmother intentionally, whereas only 79% said that he shot the bulls-eye intentionally. Id. This result is just more evidence that the Knobe Effect operates when the person foresees the consequences of her actions. 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. But see FRITZ HEIDER, THE PSYCHOLOGY OF INTERPERSONAL RELATIONS 83 (1958) (arguing that “trying” and “intending” are the same thing).

127 Nadelhoffer, supra note 4, at 345

128 Id.

129 On the other hand, one might say that committing a felony, or at least a dangerous felony, implies some level knowledge that someone’s life is being endangered. See Part II.C, supra. Additionally, Nadelhoffer conducted a follow-up 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 guesses the code to cause a bad result, people rated his action as very intentional, and when he correctly guesses the code to cause a good result, people rated his action as less intentional. Nadelhoffer, supra note 4, at 347–49. The study, however, evaluated two other dimensions: the extent to which people praised or blamed Fred for the result, and the distinction between whether he intentionally guessed the right numbers and whether he intentionally caused the ultimate result. Id. The results of Nadelhoffer’s study are as follows.
C. The Knobe Effect Represents a Mistaken Perception

Some argue that the Knobe Effect represents a mistaken perception of a person’s mental state. On this view, morally bad conduct and outcomes lead people to mistakenly assume that a person subjectively intended to cause the bad outcome, which is why they mistakenly label the action intentional. If the Knobe Effect simply represents this error, then Knobe’s theory cannot address the retributivist criticism of felony murder as I have proposed. In fact, if the Knobe Effect suggests that felony murder is based on a mistaken assumption, then it could support abolishing the rule entirely.

There are at least three ways in which the Knobe Effect may represent a mistaken perception of mental states. It may be infected by hindsight bias, motivational bias, and outcome bias. I use the term “infect” because these biases, although they contribute to the Knobe Effect, may not be its sole cause. Even so, such

---

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.

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.

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. 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.

---

problems could undermine the argument that felony murder deals with intentional killings. ¹³¹

1. Hindsight Bias

Hindsight bias, perhaps the simplest of the three, may infect both the Knobe’s theory and felony murder. Generally speaking, it is “the phenomenon that people overestimate the predictability of past events.”¹³² As it relates to the Knobe’s theory, hindsight bias may cause people to overestimate the predictability of morally bad outcomes, thus leading them to conclude more often than they should that the person subjectively intended to cause or risk the outcome.¹³³

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

¹³¹ A fourth bias that could infect the Knobe Effect 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 more general and broadly applicable than the other three biases, I have not discussed it in detail. However, as with any judgment about culpability or intent, the Actor-Observer Bias likely infects judgments in felony murder cases.


¹³³ This explanation might comport with one of the common utilitarian justifications for felony murder, which is that it deters people from committing felonies that they know will put someone’s life in danger. See, e.g. Donald Baier, Note, Arizona Felony Murder: Let the Punishment Fit the Crime, 36 ARIZ L. REV. 701, 712–13 (1994) (citing 2 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, which leads them to conclude more often than they should that the defendant recognized the risk inherent in the felony. Of course, when the defendant does not foresee the danger, the deterrence rationale makes less sense.
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.\footnote{Lijtmaer, supra note 132, at 639 (citing Rachlinski, supra note 132, at 584).}

If the Knobe Effect occurs when people mistakenly attribute foresight to people who did not have it, then it represents an error. Believing that a person foresaw the consequences of her actions, an observer could conclude that she subjectively intended them, or at least took the risk that they would occur. But if the perception of foresight is incorrect, then the conclusion about intent is incorrect as well.

Thus, hindsight bias could undermine felony murder laws. Lijtmaer explains that hindsight bias increases felony murder punishments because it “makes juries more likely to conclude that a death occurring during the course of a felony was foreseeable.”\footnote{Id at 645. Presumably, if the death was foreseeable, then the defendant is more culpable, although Lijtmaer does not make this connection.} The Knobe Effect would only exacerbate this problem by leading people to believe that the death was not only culpable, but also intentional.

On the other hand, the problem of hindsight bias might 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.”\footnote{See Part II.C, supra.} I argued that this formulation best ensures that the defendant actually had the sub-baseline intentions required by Knobe’s theory,\footnote{See Part II.C, supra.} reasoning that Binder’s formulation, which limits felony murder to predetermined lists of “inherently” dangerous felonies, was less accurate because even the most dangerous felonies do not involve sub-baseline intentions in every case.\footnote{See Part II.C, supra.} Hindsight bias, however, appears to undermine Crump’s formulation, because Crump’s formulation...
depends on people’s perceptions of each case, the very perceptions most infected by hindsight bias.

Therefore, perhaps Binder’s formulation best addresses the problem of hindsight bias. Limiting felony murder to a list of predetermined felonies would seem to take the biased perception out of the equation. For example, no matter how strongly jurors believed that an online hacker intended to cause one of his victims to have a heart attack, they could not convict him of felony murder if cybercrimes were not on the list of dangerous felonies (as one would hope). Again, to offer maximum protection to defendants, states should probably adopt both Binder’s and Crump’s limitations, requiring that the underlying felony be included on the list and that the jury perceive that it was committed in a way that was “clearly dangerous to human life.” This combination would help to counteract hindsight bias and ensure that the rule punishes only intentional killings.

2. Motivational Bias

Similar to hindsight bias, motivational bias may suggest that the Knobe Effect represents a mistaken perception of a person’s mental state, and thus that it does not add to the definition of intent. Motivational bias is the phenomenon by which people are motivated to attribute subjective intent to a person who causes a bad outcome. When observers feel that the person who caused something also deserves blame for it, they are motivated to conclude that the person subjectively intended it. If the Knobe Effect is infected by this prejudice, then it may largely be a side-effect of an unjustified inclination to blame.

139 Knobe, supra note 5, at 321–24.
140 Id. This effect has also been described in terms of the need to be seen blaming the person for the morally bad outcome. See, e.g., Knobe & Burra, supra note 1, at 119–120 (citing Adams & Steadman 2004). In other words, the Knobe Effect might be explained by the fact that people want to make sure they are not perceived as letting an actor off the hook for causing a morally bad outcome. Id. Thus, people will say the action was intentional to make sure they are perceived as properly condemning the person. Id.
141 Knobe, supra note 5, at 321–24. Studies appear to show, however, that motivational bias is not based on emotion. 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.
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.”\textsuperscript{142} After initially feeling that a person is to blame, the observer reinterprets the situation so that the person appears more blameworthy. Then, the observer feels even more strongly that the person is to blame, and therefore reinterprets the situation again so that such feelings make sense. Any one of these reinterpretations, or a few of them in sequence, could arrange the events in a way that makes it seem as though the person subjectively intended the bad outcome.\textsuperscript{143} In this way, the initial desire to blame sets off a series of interpretations that lead observers to conclude, more often than they should, that bad outcomes were caused intentionally.\textsuperscript{144}

If the Knobe Effect is caused by motivational bias, then it does not suggest that states can limit felony murder to intentional killings. The knee-jerk feeling that a person is to blame is not a

\textsuperscript{142} Knobe, \textit{supra} note 5, at 321.

\textsuperscript{143} Id.

\textsuperscript{144} As Nadelhoffer says, “judgment[s] concerning the moral blameworthiness or praiseworthiness of agents can have a similar influence on our ascriptions of intentional action.” Nadelhoffer, \textit{supra} note 130, at 205 (citing Nadelhoffer 2204b). 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, \textit{supra} note 5, at 321 (citing Ditto et al. 2009; Haidt 2001).
rational way to evaluate intent. Retributive theory requires specific indicators of culpability before authorizing punishment.\textsuperscript{145}

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.\textsuperscript{146} He cites a study involving two hypothetical situations: in one, morally bad conduct caused a morally good outcome; in the other, morally good conduct caused a morally good outcome. Because the outcomes were good, there was no need for blame. Nevertheless, people were more likely to call the action intentional when the conduct was bad.\textsuperscript{147}

This study suggests that the Knobe Effect depends on the morality of the conduct more than the outcome. As Knobe says,

\begin{quote}
[T]he 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.\textsuperscript{148}
\end{quote}

If the Knobe Effect is caused by the morality of conduct, not outcomes, then it is not infected by motivational bias, because motivational bias deals only with outcomes. Additionally, judging a person based on her conduct is consistent with Binder and

---

\textsuperscript{145} See Part II.B, supra.
\textsuperscript{146} Knobe, supra note 5, at 322 (citing Hitchcock & Knobe 2009).
\textsuperscript{147} Knobe, supra note 5, at 323.
\textsuperscript{148} Id. (emphasis removed). Indeed, if motivational bias is causing the Knobe Effect in this scenario, 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.’

Id. at 322. If such an explanation is unlikely, then perhaps the Knobe Effect operates independently of judgments regarding blame. See also the study by Guglielmo & Malle, showing that people judge whether an action is intentional \textit{before} they judge whether the actor is to blame for the outcome. Id. On the other hand, the participants could have been assigning blame for the bad conduct, regardless of the outcome.
Crump, who suggest that a person’s conduct, apart from her mental state, is relevant to her culpability.\textsuperscript{149} In fact, the theories of Binder and Crump seem well-suited to combating motivational bias, because their proposed limitations on felony murder would force factfinders to focus on conduct apart from emotionally-laden outcomes.\textsuperscript{150}

3. Outcome Bias

Outcome bias, like the other two biases, suggests that the Knobe Effect represents a mistaken perception. Outcome bias suggests that judgments about a person’s conduct are influenced by the outcomes that it causes.\textsuperscript{151} As explained in the previous section, motivational bias is not a serious problem for Knobe’s theory because conduct, not outcomes, cause people to label an action intentional.\textsuperscript{152} Outcome bias, however, suggests that morally bad outcomes make people judge conduct too harshly, thus causing them to attribute more intent to the action than they should. In this way, outcome bias may cause observers to label a killing intentional when it was not.\textsuperscript{153}

\textsuperscript{149} See Part II.B, supra. Binder, of course, argues that outcomes are relevant to culpability, Binder, supra note 61, at 1030–31, 1034, but this is not necessarily inconsistent with removing judgments about outcomes from the evaluation of culpability. See Part III.C, infra, discussing how conduct and outcomes can be prejudged in the abstract to avoid bias.

\textsuperscript{150} A problem with this argument, however, is that felony murder cannot be divorced entirely from outcomes. By definition, it depends on the fact that a death occurred. Moreover, as Binder argues, society should punish completed attempts more severely than failed attempts. Binder, supra note 61, 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, 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, which diminishes the effect of motivational bias.

\textsuperscript{151} Lijtmaer, supra note 132, at 641 (explaining that outcome bias is “the tendency to judge the quality of a decision, good or bad, based on its consequences”) (citing Philip G. Peters, Jr., Hindsight Bias and Tort Liability: Avoiding Premature Conclusions, 31 Ariz. St. L.J. 1277, 1282 (1999)).

\textsuperscript{152} Knobe, supra note 5, at 322–23; Part III.C.2, supra.

\textsuperscript{153} 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 when it causes an innocent person’s death. As Lijtmaer explains,
Outcome bias presents a serious problem for my argument that states can restrict felony murder to intentional killings. All may agree that a killing is not intentional under Knobe’s theory unless the conduct exhibited sub-baseline intentions regarding death. I argued that this was the proper limit for felony murder—in accordance with Solan, Binder, and Crump—because it ensured that the rule only applied to intentional killings. The problem, however, is that people may not be able to judge when a defendant’s conduct exhibits sub-baseline intentions, because a morally bad outcome (of which death must be the worst) biases their judgments about the conduct itself.

There is a sense, however, in which 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 behavior during felonies, demonstrate sub-baseline intentions regarding 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 a killing intentional under Knobe’s theory. 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 this bias by limiting felony murder to types of conduct that satisfy a set of predetermined conditions indicating the requisite sub-baseline intentions. Again, thirty five states use lists of “inherently” dangerous felonies for this purpose.

“[R]esearch suggests that [outcome bias] is amplified as the severity of the injury increases.” Lijtmaer, supra note 132, at 643 (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); Peters, supra note 151, at 1283 (“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 643 (citing Caplan, supra note 153, at 1960).

154 See Part II.C, supra.
and Crump’s formulation—which looks for conduct during a felony that is “clearly dangerous to human life”—could serve the same purpose.\(^{155}\) Thus, outcome bias can be avoided by deciding in advance which types of conduct support the conclusion that the defendant had the sub-baseline intentions necessary to call the killing intentional.\(^{156}\)

Three problems remain. First, as Tomkovicz and Gerber point out, politics still influences which felonies and types of conduct are included on the list, which means the list will probably be too long.\(^ {157}\) Second, outcome bias influences which felonies and types of conduct are included as well. After all, the study cited by Lijtmaer involved hypothetical scenarios, not real people, and the participants still exhibited significant levels of outcome bias.\(^ {158}\)

Third, sentencing decisions are unavoidably infected by outcome bias. Perhaps this is the most difficult problem to resolve. On the one hand, outcome bias likely infects sentencing in all cases, not only felony murder cases. On the other hand, the fact that felony murder authorizes murder-level punishments even when the defendant did not subjectively intend to kill or recklessly endanger seems to exacerbate the extent to which biases can mar a judge’s sentencing decision. Simply put, felony murder allows judges to be too harsh.\(^ {159}\) Moreover, although lawmakers can

\(^{155}\) For example, states could list certain kinds of conduct, similar to aggravating factors, that the prosecution would have to prove for a conviction on felony murder.

\(^{156}\) Thus, whatever role outcome bias plays in causing the Knobe Effect would be negated as well. Even if people wanted to attribute intent because a morally bad outcome biased their perception of Solan’s baseline, they would not be able to impose punishment for intentional murder unless the action met the enumerated set of conditions prejudged in the abstract to fall below the baseline.

\(^{157}\) See Part III.A, supra.

\(^{158}\) 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 reason for the inquiry—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 likely to overestimate the culpability of the felony in question.

\(^{159}\) 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
prejudge felonies and conduct, they cannot prejudge sentencing decisions, because determinate sentences are unconstitutional.\textsuperscript{160} Outcome bias in sentencing, therefore, merits further consideration, especially with regard to felony murder.

CONCLUSION

I have argued that Knobe’s theory provides a framework by which states can limit felony murder to intentional killings, thus answering the retributivist charge that felony murder imposes punishment without evidence of culpability. The question remains, however: Should theories like Knobe’s have any bearing on how criminal laws are designed? Should the law be structured on a rational basis, rather than bending to theories that depend on the popular understanding of certain words?

At one level, the question is beyond the scope of this article. My argument simply presumes that retributive theory should take into account what people actually mean when they talk about intent and culpability. I also presume that when psychology or linguistics discovers something about these concepts, it is likely to be relevant to the way in which a society structures its laws. Moreover, I explicitly disclaim any discussion of utilitarian criminal theory, recognizing, nonetheless, that many believe retributivism itself is outdated, brutish, or simply ineffective at accomplishing the goals of criminal law.

In the end, however, the question must be answered, and for the purposes of this article, I will answer it affirmatively. I believe it is important to discover, through a variety of disciplines, what people mean when they talk about culpability, intent, and justice—and that it is important to reflect, when discoveries are made in these areas, on how criminal laws could be reformed. The need for reflection, moreover, is evident under all theories of punishment—retributive, utilitarian, or otherwise. Thus, perhaps this article’s most important contribution is not its argument, but rather its embodiment of the principle that the criminal law should always be willing grapple with new theories and discoveries, re-


testing its justification each time. Considering how grave it is to deprive someone of liberty, I believe such reflection is worthwhile.