

# CRIMINAL LAW

## THE COGNITIVE PSYCHOLOGY OF MENS REA

KEVIN JON HELLER\*

*This Article provides a comprehensive—though admittedly speculative—explanation of how jurors use projection and prototyping, two simulation-based theories of mindreading, to make mental-state attributions in criminal cases. The first two Parts explain why jurors are unlikely to use a functionalist method of mindreading in a case that focuses on the defendant’s mens rea. The next three Parts introduce projection and prototyping, describe the evidence that jurors actually use them to make mental-state determinations, and discuss the cognitive mechanism—perceived similarity between juror and defendant—that determines which one a juror will use in a particular case. The final two Parts explain why projection and prototyping are likely to result in inaccurate mental-state determinations and discuss debiasing techniques that may make them more accurate.*

### I. INTRODUCTION

*Actus non facit reum nisi mens sit rea*—“the act does not make a person guilty unless the mind is also guilty.”<sup>1</sup> Few today would disagree with the maxim; the criminal law has long since rejected the idea that causing harm should be criminal regardless of the defendant’s subjective culpability.<sup>2</sup> Still, the maxim begs a critical question: can jurors accurately

---

\* Senior Lecturer, Melbourne Law School. J.D. 1996, Stanford Law School; M.A. 1993, Duke University; M.A. 1991, New School for Social Research. Thanks are due to Scott Optican, John Ip, Tommy Crocker, Alice Ristroph, Bruce Hay, and Dan Kahan for their comments on earlier drafts of this Article, as well as to Kitaj Woodward for her excellent research assistance.

<sup>1</sup> WAYNE R. LAFAVE, CRIMINAL LAW § 3.4, at 225 (4th ed. 2003).

<sup>2</sup> See, e.g., Martin R. Gardner, *The Mens Rea Enigma: Observations on the Role of Motive in the Criminal Law Past and Present*, 1993 UTAH L. REV. 635, 642-43 (“[C]ommentators generally agree that no systematic mens rea requirements existed until at least the early thirteenth century. Indeed, prior to that time it appears the law was essentially grounded in strict liability.”).

determine whether the defendant acted with the requisite “guilty mind”? St. Thomas Aquinas was certainly skeptical that such mindreading—as cognitive psychologists call it<sup>3</sup>—is within the ken of mere mortals:

[M]an, the framer of human law, is able to judge only of outward acts; because *man seeth those things that appear*, according to 1 Kings 16.7; but God alone, the framer of the Divine law, is able to judge the inward movement of wills . . . .<sup>4</sup>

Given the significant cognitive demands contemporary criminal law imposes on jurors, Aquinas’s skepticism seems more prescient than ever. The common law, for example, not only asks jurors to distinguish between seventy-eight different terms for mental states,<sup>5</sup> it fails to define the various mental states consistently—there are seven different definitions of *willful* alone<sup>6</sup>—and combines them in permutations that defy comprehension.<sup>7</sup> Indeed, the common law is such a mess that no less an authority than Justice Robert Jackson once bemoaned the “variety, disparity, and confusion” of the “requisite but elusive mental element.”<sup>8</sup>

The Model Penal Code (MPC), adopted by a majority of states,<sup>9</sup> may actually be worse. To be sure, the MPC takes a far more systematic approach to mens rea, winnowing the common law’s seventy-eight mental states to four: purpose, knowledge, recklessness, and negligence.<sup>10</sup> Those mental states are differentiated with such “subtlety and precision,” however, that it is an open question whether jurors can accurately distinguish them.<sup>11</sup>

---

<sup>3</sup> See Alvin Goldman & Kelby Mason, *Simulation*, in HANDBOOK OF THE PHILOSOPHY OF SCIENCE: PHILOSOPHY OF PSYCHOLOGY AND COGNITIVE SCIENCE 267, 267 (Paul Thagard ed., 2006).

<sup>4</sup> II ST. THOMAS AQUINAS, *SUMMA THEOLOGICA*, in 20 GREAT BOOKS OF THE WESTERN WORLD, second part, part I, question 100, art. 9, at 261 (Robert Maynard Hutchins ed., founders ed. 1952). Modern legal scholars have expressed similar skepticism. See, e.g., H.L.A. HART, *PUNISHMENT AND RESPONSIBILITY* 33 (1968) (noting that although “[i]t is relatively simple to show that an agent lacked . . . volitional muscular control[,] it is somewhat more difficult to show that he did not know certain facts about either present circumstances . . . or the future” and “much more difficult to establish whether or not a person was deprived of ‘self-control’ by passion provoked by others, or by partial mental disease”).

<sup>5</sup> See, e.g., Paul H. Robinson & Jane A. Grall, *Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond*, 35 STAN. L. REV. 681, 692 n.48 (1983).

<sup>6</sup> *Id.*

<sup>7</sup> See, e.g., Paul H. Robinson et al., *The Five Worst (and Five Best) American Criminal Codes*, 95 NW. U. L. REV. 1, 29 (2000) (noting that common-law criminal codes “frequently use confusing, convoluted, or arcane language in setting out offenses”). Rhode Island, for example, defines murder as a “willful, deliberate, malicious, and premeditated killing.” R.I. GEN. LAWS § 11-23-1 (2002 & Supp. 2008).

<sup>8</sup> MODEL PENAL CODE & COMMENTARIES § 2.02 cmt. at 230 (1985).

<sup>9</sup> See Robinson & Grall, *supra* note 5, at 691-92.

<sup>10</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(2).

<sup>11</sup> Rebecca Dresser, *Culpability and Other Minds*, 2 S. CAL. INTERDISC. L.J. 41, 48

Consider, for example, the MPC's three subjective mental states: a person acts "purposely" if his "conscious object" is to bring about a particular result;<sup>12</sup> acts "knowingly" if he is "aware that it is practically certain" that his conduct will lead to the result;<sup>13</sup> and acts "recklessly" if he "consciously disregards a substantial and unjustifiable risk" that his conduct will cause the result.<sup>14</sup> Those are fine distinctions, to say the least. Little wonder, then, that scholars have described the MPC as an "elaborate set of precise rules whose operability depends on the jury's willingness"—to say nothing of their ability—"to make artificial characterizations."<sup>15</sup>

The MPC's idiosyncratic definition of negligence only further complicates mindreading. Negligence is defined as the defendant acting "when he should be aware of a substantial and unjustifiable risk" that his conduct will lead to a particular result.<sup>16</sup> Common law negligence is not actually a mental state; the reasonable-person standard "is determined and applied without reference to what the actor was thinking at the moment."<sup>17</sup> The MPC, by contrast, subjectivizes negligence, once again requiring jurors to read the defendant's mind: whether the defendant's conduct was unreasonable must be determined "considering the nature and purpose of his conduct and the circumstances known to him" at the time of the crime.<sup>18</sup>

Nor is that all. Although the common law at least limited each offense to a single mental state,<sup>19</sup> the MPC permits different mental states to apply to different material elements of an offense—what is known as "element analysis."<sup>20</sup> Misdemeanor indecent exposure is an example: "A person commits a misdemeanor if, for the *purpose* of arousing or gratifying sexual desire . . . he exposes his genitals under circumstances in which he *knows* his conduct is likely to cause affront or alarm."<sup>21</sup> Thus defined, a jury must

---

(1993).

<sup>12</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(2)(a).

<sup>13</sup> *Id.* § 2.02(2)(b).

<sup>14</sup> *Id.* § 2.02(2)(c).

<sup>15</sup> Harold Edgar, *Mens Rea*, in ENCYCLOPEDIA OF CRIME AND JUSTICE 1028, 1039 (Sanford H. Kadish ed., 1983) (asserting that the Model Penal Code's (MPC) mental states are "artificial characterizations" and questioning whether they are "provable occurrences").

<sup>16</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(2)(d).

<sup>17</sup> *See, e.g.*, Adam Candeub, Comment, *Motive Crimes and Other Minds*, 142 U. PA. L. REV. 2071, 2075 n.24 (1994) (citing W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 32, at 173-74 (5th ed. 1984)).

<sup>18</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(2)(d). This injection of subjectivity into an otherwise objective test is typical of the Model Penal Code. As Gardner has pointed out, the MPC's "preference for subjective culpability is manifested throughout the substantive provisions of the Code." Gardner, *supra* note 2, at 684.

<sup>19</sup> Robinson & Grall, *supra* note 5, at 683.

<sup>20</sup> *Id.* at 703.

<sup>21</sup> *Id.* at 699 (citation omitted).

determine both whether the defendant acted *knowingly* (with regard to his conduct's potential to cause affront or alarm) and whether he acted *purposely* (with regard to arousing or gratifying sexual desire)—a far more complicated mindreading task than determining a single mental state.

This discussion also assumes that jurors normally determine whether the defendant is guilty of a single crime. In many cases, of course, jurors will have to choose between *multiple* closely-related crimes, whether because the prosecution has charged them in the alternative or because they are added in jury instructions as lesser-included offenses. Jurors in a homicide case, for example, may be asked to determine whether the defendant is guilty of first-degree murder, second-degree murder, reckless homicide, or criminally negligent homicide. Those charges are normally distinguished solely by the mental state they require: a “premeditated and intentional” killing for first-degree murder;<sup>22</sup> a “knowing” killing for second-degree murder;<sup>23</sup> a “reckless” killing for reckless homicide;<sup>24</sup> and “criminally negligent conduct” that results in death for criminally negligent homicide.<sup>25</sup>

Finally, in most criminal cases, jurors will have to deal with yet another mental state: namely, the one associated with the defense the defendant uses to negate the mens rea of the charged crime.<sup>26</sup> A defendant may defend himself against a murder charge, for example, by alleging that he was legally insane at the time of the killing.<sup>27</sup> Most often, as in the murder example, the focus will be on the defendant's lack of a “guilty mind.” In some cases, however, the defendant may invoke a defense that shifts attention to the *victim's* mental state at the time of the crime—a rape case, for example, in which the defendant claims that the victim consented to the sex.

It is fair to say, in short, that contemporary criminal law requires jurors to be latter-day Kreskins—to not only reliably distinguish nearly-indistinguishable mental states, but also to accurately determine which of many possible mental states the defendant actually possessed at the time of

---

<sup>22</sup> TENN. CODE ANN. § 39-13-202(1) (2006 & Supp. 2008).

<sup>23</sup> TENN. CODE ANN. § 39-13-210(1) (2006).

<sup>24</sup> TENN. CODE ANN. § 39-13-215(a) (2006).

<sup>25</sup> TENN. CODE ANN. § 39-13-212(a) (2006).

<sup>26</sup> See, e.g., Dresser, *supra* note 11, at 51 (“[M]any criminal defenses require mental state determinations. Duress, self-defense, intoxication, insanity, and diminished capacity are examples.”).

<sup>27</sup> See, e.g., JENNY MCEWAN, *THE VERDICT OF THE COURT* 75 (2003) (“[T]he underlying principle of the insanity defense . . . amounts to a denial of *mens rea*.”).

the crime.<sup>28</sup> Is such mindreading possible? Or is Aquinas correct that “God alone . . . is able to judge the inward movement of wills”?

Given the centrality of mens rea to criminal responsibility, we would expect legal scholars to have provided a persuasive answer to this question. Unfortunately, nothing could be further from the truth. Most scholars simply *presume* that jurors can mindread accurately, and those that take mindreading seriously have uniformly adopted “common sense functionalism,”<sup>29</sup> a theory of mental-state attribution that is contradicted by a vast amount of research into the cognitive psychology of mindreading. Common-sense functionalism assumes that a juror can accurately determine a defendant’s mental state through common-sense generalizations about how external circumstances, mental states, and physical behavior are causally related.<sup>30</sup> Research indicates, however, that mindreading is actually a *simulation*-based, not *theory*-based, process. When a juror perceives the defendant to be similar to himself, he will mindread through “projection,” attributing to the defendant the mental state that he would have had in the defendant’s situation. And when the juror perceives the defendant to be dissimilar to himself, he will mindread through “prototyping,” inferring the defendant’s mental state from the degree of correspondence between the defendant’s act and his pre-existing conception of what the “typical” crime or defense of that type looks like.

The goal of this Article is to provide a comprehensive—though admittedly speculative—explanation of how jurors use projection and prototyping to make mental-state attributions in criminal cases. The Article is divided into six Parts. The first two provide the necessary background: Part II sketches the traditional functionalist explanation of mindreading, and Part III explains why jurors are unlikely to use a functionalist method in a case that focuses on the defendant’s mens rea. The next two Parts are descriptive: Part IV introduces projection and prototyping and discusses the evidence that jurors use them to make mental-state determinations, while Part V explains the cognitive mechanism—perceived similarity between juror and defendant—that determines which technique a juror will use in a

---

<sup>28</sup> See, e.g., NORMAN J. FINKEL, COMMONSENSE JUSTICE 61-62 (1995) (“We place a great burden on jurors . . . [W]e ask jurors to comprehend the act, infer the intent, and reach a judgment about culpability and blameworthiness. Particularly in terms of fathoming *intent*, jurors must plunge from the hazy facts into a defendant’s mind, where a darkness that eludes even videotapes reigns.”); Dresser, *supra* note 11, at 48 (“With the current law’s increased emphasis on subjective culpability, other minds judgments have become more imperative, because the legitimacy of the subjective approach absolutely depends on our ability to determine whether defendants possessed the requisite state of awareness when the criminal activity occurred.”).

<sup>29</sup> Goldman & Mason, *supra* note 3, at 269.

<sup>30</sup> See *infra* notes 38-53 and accompanying text.

particular case. The final two Parts are then analytic: Part VI explains why projection and prototyping are likely to result in inaccurate mental-state determinations, and Part VII discusses de-biasing techniques that could be used to improve their accuracy.

## II. THE TRADITIONAL EXPLANATION

Rebecca Dresser has argued that “the general legal reaction to the proof issues raised by other minds judgments is either to ignore or to dismiss them.”<sup>31</sup> It is difficult to disagree. Aside from Dresser’s own article and an article by Adam Candeub,<sup>32</sup> one searches the legal literature in vain for a sustained discussion of the psychology of juror mindreading. Most scholars, it appears, simply *presume* that accurate mindreading is possible—a position that dates back at least to 1882, when Bowen famously dismissed Oliver Wendell Holmes’s skepticism toward mindreading<sup>33</sup> by claiming that “the state of a man’s mind is as much a fact as the state of his digestion.”<sup>34</sup> Little has changed in the intervening century. In fact, as recently as 1987, Richard Singer could assert without argument that “[i]t is certainly within the jury’s ken to find that a typical self defender did not intend (purpose) a killing.”<sup>35</sup>

The question, of course, is *why* legal scholars take it for granted that jurors can make accurate mental-state determinations. As discussed above,<sup>36</sup> given the significant cognitive demands that contemporary criminal law imposes on jurors, it is far from obvious that they can. The answer seems to be that legal scholars embrace, implicitly or explicitly, a commonsense theory of mental-state attribution in which mindreading seems neither particularly complicated nor particularly problematic.

That theory relies on three interrelated assumptions. The first is that there are fixed and unchanging causal relationships between external circumstances, internal mental states, and physical behavior. Jerome Hall’s response to Holmes—echoing Bowen—is illustrative:

[T]he inevitable limitations on our knowledge do not support the conclusion that, in the vast majority of judgments based on rational methods of investigation, there is no

---

<sup>31</sup> Dresser, *supra* note 11, at 61.

<sup>32</sup> See generally Candeub, *supra* note 17.

<sup>33</sup> See Jerome Hall, *Interrelations of Criminal Law and Torts: I*, 43 COLUM. L. REV. 753, 769 (1943) (noting that Holmes “saw fit to take the road of rather extreme skepticism, because he apparently believed that there must be a great disparity between conclusion of fact based on external conduct and actual inner states”).

<sup>34</sup> *Id.* at 767 (citation omitted).

<sup>35</sup> Richard Singer, *The Resurgence of Mens Rea: II—Honest but Unreasonable Mistake of Fact in Self Defense*, 28 B.C. L. REV. 459, 515 (1987).

<sup>36</sup> See *supra* notes 6-28 and accompanying text.

reasonably accurate correspondence. The whole law of evidence is a studied effort to cope with this fundamental problem and its justification rests on the high probability of a “sufficiently” accurate representation of inner states by external conduct. This premise is accepted not only in law, but throughout every avenue of social life. It rests ultimately on the essential uniformity of human nature, and is so deeply ingrained in our daily experience that it is hardly possible even to conceive of social intercourse founded on its rejection.<sup>37</sup>

The second assumption, dependent on the first, is that jurors possess an *intuitive theory* of the specific causal relationships that exist between circumstances, mental states, and behavior. Wigmore divides those relationships—what H.L.A. Hart calls “common-sense generalizations about human nature”<sup>38</sup>—into two basic categories: (1) “circumstances tending to excite, stimulate, or bring the emotion in question into play”; and (2) “outward conduct expressing and resulting from the emotion in question.”<sup>39</sup> Hart’s insistence “that men are capable of self-control when confronted with an open till but not when confronted with a wife in adultery”<sup>40</sup> and Oliver Wendell Holmes’s belief that “[d]etached reflection cannot be demanded in the presence of an uplifted knife”<sup>41</sup> are examples of commonsense generalizations in the first category. *Rex v. Shaw*’s assertion that strangling someone to death with a rope indicates malice aforethought is an example of a commonsense generalization in the second category.<sup>42</sup> As the examples indicate, both categories of generalizations take a common “if *x*, then *y*” form, where *x* is the circumstances or behavior and *y* is the mental state that can be inferred from them.<sup>43</sup>

The third and final assumption then flows naturally from the second: jurors determine a defendant’s mental state by applying their commonsense theory of the specific relationships that exist between circumstances, mental states, and behavior to the evidence in the case. Hall makes that point explicitly in his discussion of how fact-finders rely on “the essential

---

<sup>37</sup> Hall, *supra* note 33, at 770-71.

<sup>38</sup> HART, *supra* note 4, at 33.

<sup>39</sup> 2 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 387, at 416 (1979).

<sup>40</sup> HART, *supra* note 4, at 33.

<sup>41</sup> *Brown v. United States*, 256 U.S. 335, 343 (1921).

<sup>42</sup> (1834) 172 Eng. Rep. 1282 (N.P.); see also Luke Wilson, *Renaissance Tool Abuse and the Legal History of the Sudden*, in LITERATURE, POLITICS, AND LAW IN RENAISSANCE ENGLAND 121, 128-39 (Erica Sheen & Lorna Hutson eds., 2005) (noting that courts in the Renaissance era tended to infer lack of premeditation from “tool abuse”—killing with an object that was not designed to serve as a weapon).

<sup>43</sup> Philosophers and cognitive psychologists who adopt functionalist theories tend to rely on more sophisticated and less concrete generalizations, such as “[i]f someone wants *x*, and believes that *y* is a means to *x*, then, other things being equal, they will do *y*.” Goldman & Mason, *supra* note 3, at 268. The basic idea, however, is the same.

uniformity of human nature” to determine what kinds of behavior indicate that a defendant acted “intentionally”:

[H]uman conduct that is associated causally with certain harms proscribed by law, is labelled [sic] “intentional” by triers of the material facts on the basis of knowledge of certain external data . . . . Given certain facts, we *must* conclude that any and every rational human being in those circumstances did or did not *intend* the results . . . .<sup>44</sup>

Judge Posner offers a more specific example of this reasoning process—what he calls, somewhat idiosyncratically, a “behaviorist account of deliberation”<sup>45</sup>—when he discusses how jurors infer premeditation from the evidence in a case:

[I]n deciding whether a crime is premeditated . . . we examine the circumstances of the crime: Was it concealed? Had the criminal made arrangements for a getaway? Had he obtained the means of committing the crime in advance? Were those means suitable to the end (suitably lethal, in the case of a murder)? Did the criminal have much to gain from the crime? From these circumstances a model of a deliberating criminal could be constructed—an “objective” reconstruction of the criminal’s motivational experience, created by attributing to him a certain type of rationality.<sup>46</sup>

Taken together, these three assumptions constitute what cognitive psychologists have described as a functionalist account of mindreading.<sup>47</sup> We can summarize that account as follows:

[O]ur mindreading capacity is implemented by an intuitive theory of mind, a body of knowledge or belief about the causal relations between mental states on the one hand and behavior, environment and other mental states on the other . . . . In mindreading, we use these generalizations and our data about the target’s circumstances or behavior to infer some of the target’s mental states.<sup>48</sup>

Two additional aspects of functionalist theorizing are worth noting. First, the theory assumes that jurors acquire their “intuitive theory of mind” in a manner that approximates the scientific method: via “a gradual process of theorizing, making generalizations based on the available evidence, revising . . . theories to handle incongruent evidence, etc.”<sup>49</sup> The process, in other words, is *empirical*,<sup>50</sup> grounded in jurors’ real-world observation of

---

<sup>44</sup> Hall, *supra* note 33, at 767; *see also* GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART 91 (2d ed. 1961) (arguing that “[i]t is platitudinous to say that intent must frequently be gathered from conduct on the basis of a supposed uniformity of human nature”).

<sup>45</sup> RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 175 (1990).

<sup>46</sup> *Id.*

<sup>47</sup> Goldman & Mason, *supra* note 3, at 269.

<sup>48</sup> *Id.* at 286-89.

<sup>49</sup> *Id.* at 270.

<sup>50</sup> ALVIN I. GOLDMAN, SIMULATING MINDS 26 (2006); *cf.* WIGMORE, *supra* note 39, § 394, at 437 (“Drawing inferences from . . . conduct always proceeds from experience.”).

causal relationships between external circumstances and mental states and between mental states and physical behavior.<sup>51</sup>

Second, as conceptualized by functionalist theorizing, mindreading does not involve any kind of introspection on the part of the juror—the juror mindreads by applying his commonsense generalizations to the evidence in the defendant’s case; he does not ask what mental state *he* would have had in the defendant’s situation.<sup>52</sup> The distinction is a fine one: given that the juror’s generalizations are ostensibly based on the psychology of “any and every rational human being,” they should apply no less to him than to others. Nevertheless, although theorizing and introspection may well lead a juror to attribute the same mental state to the defendant, the *processes* they employ are completely different: functionalism views mindreading as the application of a theory, not as an exercise in phenomenology.

### III. PROBLEMS WITH THE TRADITIONAL EXPLANATION

There is no question that functionalist theorizing provides an intuitively plausible explanation of how jurors mindread.<sup>53</sup> Jurors do not have direct access to the defendant’s mind, but they do know what the defendant’s physical behavior was—the crime itself, the actus reus—and have at least some sense of what the external circumstances were that led to that behavior. It thus seems eminently sensible to assume that a juror would infer the “missing” mental state from the evidence by relying on an intuitive theory of how circumstances, mental states, and behavior are causally related.

For a number of reasons, however, projection and prototyping provide a far better account of how jurors make mental-state determinations than functionalist theorizing. To begin with, research indicates that the “default starting point”<sup>54</sup> of mindreading is projection or prototyping, not functionalist theorizing. In terms of prototyping, for example, Fiske and Neuberg have found that “[p]erceivers attempt category-based impression

---

<sup>51</sup> See, e.g., Dresser, *supra* note 11, at 75 (“Like other empirical generalizations, it is justified by our prior observation of a consistent connection between two phenomena—the external signs and the mental states they represent.”).

<sup>52</sup> See GOLDMAN, *supra* note 50, at 40 (“Under the [functionalist] account, the mind-reader never ascribes one of her own decisions to the target; no such decision is part of the theorizing routine. In the theorizing routine, all states of the mind-reader are (third-person) *metarepresentations*—that is, beliefs about mental states of the target.”).

<sup>53</sup> See Dresser, *supra* note 11, at 75 (“[T]his theory identifies mental states according to a process that sounds very much like what goes on in courtrooms every day.”).

<sup>54</sup> Daniel R. Ames, *Everyday Solutions to the Problem of Other Minds*, in *OTHER MINDS: HOW HUMANS BRIDGE THE DIVIDE BETWEEN SELF AND OTHERS* 158, 166 (Bertram F. Malle & Sara D. Hodges eds., 2005).

formation before they use more attribute-oriented impression formation, and if relatively category-oriented processes are successful, then the perceiver goes no further toward more attribute-oriented processes.”<sup>55</sup> Similarly, in terms of projection, Krueger has shown that “[w]hen the responses of others are not known, people project their own as a first bet.”<sup>56</sup>

Krueger’s conclusion indicates, of course, that jurors may shift from projection to functionalist theorizing if the “responses of others” are known. Indeed, Ames argues that, in general, “[c]umulative behavioral evidence supercedes extra-target strategies: [p]rojection and stereotyping will drive mindreading when behavioral evidence is ambiguous, but as apparent evidence accumulates, inductive judgments will dominate.”<sup>57</sup> That makes sense—if a mindreading target’s behavior is unambiguous because we know how the target normally reacts to certain situations or because one and only one mental state is consistent with the behavioral evidence, there is no reason to disregard that knowledge in favor of “extra-target strategies” like projection and prototyping.

Few criminal cases that focus on mens rea, however, will provide jurors with sufficiently unambiguous behavioral evidence. First, voir dire will ensure that the defendant is a stranger to the jurors,<sup>58</sup> while the character evidence rule will normally prevent the prosecution from compensating for that fact by bringing information about the defendant’s character to their attention.<sup>59</sup> Jurors will thus usually have no information concerning how the defendant normally responds to an ordinary situation, much less to an extraordinary situation like the one at issue in the case.

Second, it is very unlikely that the behavioral evidence in the case—the actus reus—will be consistent with one and only one mental state.

---

<sup>55</sup> Susan T. Fiske & Steven L. Neuberg, *A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation*, 23 *ADVANCES IN EXPERIMENTAL SOC. PSYCHOL.* 1, 2 (1990).

<sup>56</sup> Joachim I. Krueger, *Return of the Ego—Self-Referent Information as a Filter for Social Prediction: Comment on Karniol (2003)*, 110 *PSYCHOL. REV.* 585, 589 (2003).

<sup>57</sup> Ames, *supra* note 54, at 166. Note that Ames uses the term *stereotyping* instead of *prototyping*. In this context, the two are equivalent. See Jennifer Eno Loudon & Jennifer L. Skeem, *Constructing Insanity: Jurors’ Prototypes, Attitudes, and Legal Decision-Making*, 25 *BEHAV. SCI. & L.* 449, 451 (2007) (“[T]he most widely used definition of stereotypes portrays them as prototypes that focus on people and are associated with strong attitudes that are resistant to change. Thus, stereotypes can be viewed as a subtype of prototypes that are distinguished primarily by their emotional connotation.”). I use *prototyping* for consistency and clarity.

<sup>58</sup> See, e.g., WAYNE R. LAFAVE ET AL., *CRIMINAL PROCEDURE* § 22.3, at 1043 (3d ed. 2000).

<sup>59</sup> See FED. R. EVID. 404(a) (prohibiting the use of evidence of character to prove conduct in conformity therewith). “Character is a generalized description of one’s disposition . . .” *Id.* 406, advisory committee’s note (citation omitted).

Contrary to functionalist theorizing, “there is no one-to-one correspondence between events and other people’s reactions to these events; people do not always react the same way to a given event, and the same psychological reaction can be elicited in many different ways.”<sup>60</sup> That does not mean a juror will see all of the possible mental-state inferences in a criminal case as equally reasonable; the behavioral evidence may make some mental-state inferences more plausible than others. But it seems very unlikely that a juror trying to determine the defendant’s mental state in a case involving a murder or a rape, for example, will view the behavioral evidence as *unambiguous*, given that such complex acts are particularly difficult to mindread.<sup>61</sup> Moreover, the juror will not be mindreading in a vacuum: his determination will be framed by the competing descriptions of the defendant’s mental state offered by the prosecution and the defense, a conflict that cannot help but reinforce the ambiguity of the behavioral evidence.<sup>62</sup>

Both the default primacy of extra-target mindreading strategies and the general ambiguity of the *actus reus* in a criminal case indicate, in short, that jurors are likely to use projection and prototyping, instead of functionalist theorizing, to determine a defendant’s mental state. Indeed, that is exactly what a significant number of empirical studies of mindreading in the legal context have found. That research is discussed in the next Part.

#### IV. PROJECTION

##### A. PROJECTION DESCRIBED

As we have seen, the functionalist model of mindreading makes two basic assumptions: (1) that jurors have an intuitive theory of mind that consists of commonsense generalizations about the causal relationships between external circumstances, mental states, and physical behavior; and (2) that jurors mindread by applying their intuitive theory of mind to the

---

<sup>60</sup> Rachel Karniol, *Egocentrism Versus Protocentrism: The Status of Self in Social Prediction*, 110 *PSYCHOL. REV.* 564, 564 (2003); see also THOMAS O. BUFORD, *ESSAYS ON OTHER MINDS* xi (1970) (“Mental events and physical behavior are contingently related . . .”).

<sup>61</sup> See, e.g., Candeub, *supra* note 17, at 2102 (“As soon as one develops more complex acts . . . certainty falls away and one can eliminate other potentially attributable mental states far less easily. Some types of acts can be done with numerous purposes in mind, and simply observing the act or describing it in purely physical terms will not reveal which of these purposes is, in fact, controlling the actor’s mind when the act is being committed.”).

<sup>62</sup> This framing effect should be even more pronounced in cases in which the prosecution charges the defendant with related crimes that are distinguished from each other only by the mental state they require. If the prosecution does not believe that behavioral evidence is unambiguous, how can a juror?

evidence in the defendant's case and then imputing the resulting mental state to the defendant.

The projection model of mindreading conceptualizes mental-state determinations very differently. It does not assume that jurors possess an intuitive theory of mind; indeed, interest in the model was originally sparked by suspicion that functionalist mindreading "is *too* information rich—that it attributes too much knowledge and sophistication to mindreaders, even if the knowledge is only tacit or unconscious."<sup>63</sup> Instead, it assumes that individuals are "information poor" but possess a "special skill, namely, constructing pretend, simulated or imaginary mental states . . . in simulated processing."<sup>64</sup> That skill, according to the projection model, is at the heart of mindreading:

In using the simulation heuristic to detect a target's mental state . . . an attributor begins by taking the target's "perspective." That is, he *pretends* to be in certain states the target is in, as suggested by his previous information about the target . . . . He feeds these "pretend" starting states into an appropriate cognitive mechanism of his own . . . and lets it operate on them. This mechanism outputs a new state . . . and the attributor then attributes that decision to the target. In other words, the attributor tries to make his own mind "emulate" the mental sequence the target will go through . . . . The heart of this procedure is that the attributor tries to *reproduce* or *match* what transpires in the target . . . .<sup>65</sup>

This form of projection is *prospective*: the mindreader uses simulation to move forward from information about the target's situation—his physical environment and whatever is known about his psychological characteristics—to the target's mental state.<sup>66</sup> Projection can also be *retrospective*, using simulation to move backward from the target's physical behavior—the criminal act itself—to the mental state that caused it. Goldman describes this use of projection as a "generate and test strategy":

The attributor begins with a known effect of a sought-after state, often an observable piece of behavior. He generates one or more hypotheses about the prior mental state or combination of states that might be responsible for this effect. He then "tests" (one or more of) these hypotheses by pretending to be in these states, feeding them into an appropriate psychological mechanism, and seeing whether the output matches the observed evidence. When a match is found (perhaps the first match, or the "best" match), he attributes the hypothesized state or combination of states to the target.<sup>67</sup>

---

<sup>63</sup> Goldman & Mason, *supra* note 3, at 272.

<sup>64</sup> *Id.*

<sup>65</sup> Alvin I. Goldman, *Desire, Intention, and the Simulation Theory*, in *INTENTIONS AND INTENTIONALITY: FOUNDATIONS OF SOCIAL COGNITION* 207, 217 (Bertram F. Malle et al. eds., 2001).

<sup>66</sup> GOLDMAN, *supra* note 50, at 45.

<sup>67</sup> *Id.*

Neither form of projection, it is important to note, is completely atheoretical. In particular, “there might be theory in deciding which pretend inputs to construct, and in transferring out simulated state or behavior from ourselves to the target.”<sup>68</sup> The former moment is particularly important, because the goal of projection is not to determine the mental state *the mindreader* would have had in the target’s situation, but to determine what mental state the mindreader would have had in the target’s situation *if he were the target*. Accurate projection thus depends, as explored in detail below,<sup>69</sup> on the mindreader’s ability to adopt the target’s characteristics and to prevent his own characteristics from influencing the simulation process.

Despite these theoretical moments, however, the *core* of projection remains the mindreader’s imaginative act of simulating, experiencing, and reacting to the target’s situation: “acknowledging that accuracy of inputs requires informational guidance doesn’t undercut the simulational aspect of the cognitive performances in question; the inputs are inputs *for simulation*.”<sup>70</sup> Indeed, the act of projection could not be more different than functionalist theorizing, given the latter’s emphasis on the logical and mechanical application of commonsense generalizations about the causal relationships between external circumstances, mental states, and physical behavior. In particular, there is no room in functionalist theorizing for projection’s transfer of one of the mindreader’s mental states—the mental state generated by the act of simulation—onto the target; by definition, the commonsense generalizations that the functionalist mindreader applies are derived from his experientially acquired intuitive theory of mind, *not* from introspection.

If jurors use projection to mindread, they most likely rely on a hybrid form of projection, one that is *both* prospective and retrospective. In criminal cases that focus on mens rea, jurors will usually have two different kinds of information: (1) evidence concerning the defendant’s situation, such as the circumstances in which he committed the crime and the mental state that he was in prior to finding himself in those circumstances; and (2)

---

<sup>68</sup> Goldman & Mason, *supra* note 3, at 273; *see also* Nicholas Epley et al., *Perspective Taking as Egocentric Anchoring and Adjustment*, 87 J. PERSONALITY & SOC. PSYCHOL. 327, 338 (2004) (noting that adjustments to mindreaders’ simulations are “likely guided by their theories about how different perspectives and psychological states influence judgment and perception”).

<sup>69</sup> *See infra* notes 177-232 and accompanying text.

<sup>70</sup> GOLDMAN, *supra* note 50, at 175; *see also* Robert M. Gordon, *The Simulation Theory: Objections and Misconceptions*, 7 MIND & LANGUAGE 11, 18 (1992) (“[E]ven where we do make use of generalizations that correlate external circumstances with behavior, it is not the generalizations themselves, that is, the generalizations mechanically applied, that explain our predictive or explanatory competence: it is our skill at using the generalizations as heuristics or rules of thumb as we simulate others.”).

a critical “observable piece” of the defendant’s behavior, namely, the *actus reus* of the crime. They will thus have the raw material they need to project both prospectively (from the defendant’s situation to his mental state) and retrospectively (from the *actus reus* to the defendant’s mental state).

There is, moreover, another critical difference between general mindreading and juror mindreading that we need to consider. Whereas mindreaders usually have to generate the hypotheses they test through projection *on their own*, the prosecution and defense in a criminal case generate those hypotheses *for them*. In some cases, there will be only two competing hypotheses—a murder case, for example, in which the only issue is whether the defendant was sane or insane at the time of the crime. In others, there may be far more than two, such as a murder case in which the prosecution charges the defendant with various degrees of murder and the defense argues both self-defense and provocation.

What, then, does projection-based mindreading look like in the legal context? The mechanics of projection and the nature of criminal trials suggest a model in which a juror uses a two-stage process of projection to test the various mental states offered by the prosecution and defense. In stage one, the juror runs a *retrospective* simulation in which he tries to imagine having a particular mental state in the defendant’s situation. If he cannot imagine having that mental state, he rejects it and repeats the first stage with the next one. But if he can imagine having it, he then proceeds to stage two and runs a *prospective* simulation in which he tries to imagine that mental state leading him to commit the *actus reus* of the crime. If he cannot imagine committing the crime, he rejects the mental state and returns to stage one with the next possible mental state. But if he can imagine committing it, the projection process stops and he imputes that mental state to the defendant.

Variations on this model are certainly possible. Perhaps, instead of ending the projection process as soon as he finds a matching mental state, a juror tries to simulate all of the mental states offered by the prosecution and defense and chooses the state that best matches the *actus reus* of the crime.<sup>71</sup> Or perhaps projection is more of a gestalt process in which the juror simply imagines committing the crime in the defendant’s situation, identifies what mental state he is in, and imputes that state to the defendant. The critical point is that, whatever form it actually takes, projection provides a coherent explanation of how jurors mindread.

---

<sup>71</sup> See GOLDMAN, *supra* note 50, at 45.

## B. EVIDENCE OF PROJECTION

A significant amount of empirical research suggests that jurors use projection to make mental-state determinations. That research can be divided into two categories: general mindreading studies and studies of jury decision-making.

*1. General Research*

Cognitive psychologists have long recognized that projection is one of the most basic—and most stubborn—mindreading mechanisms:

When people predict the thoughts, feelings, or behaviors of others, they tend to assume that these others think, feel, and behave as they themselves do . . . . Although the strength of projection varies, no particular person characteristic or type of judgment item consistently fails to show projection. People project even when they are asked not to or when they receive feedback on the accuracy of their predictions . . . ; they project regardless of their level of cognitive busyness . . . and regardless of information they have about other individuals . . . .<sup>72</sup>

Indeed, the tendency to assume that others “think, feel, and behave” as they do is so tenacious that it routinely leads individuals to mindread *inaccurately*. As noted above, accurate projection requires the mindreader to quarantine his self-perspective: the goal of the simulation process is to determine the *target’s* mental state, not the *mindreader’s*. Nevertheless, such quarantining often fails, contaminating the simulation process. For example, “people in one emotional situation . . . project their current preferences and behaviors onto other people who are in different emotional situations”<sup>73</sup>—a phenomenon known as an “empathy gap.”<sup>74</sup> Such empathy gaps were evident in a study that asked individuals to predict whether hikers lost in the mountains without supplies would feel greater regret about not bringing food or not bringing water with them. In comparison to participants who were asked to exercise before making their predictions—and who were thus in the same “hot” emotional state as the imaginary hikers—participants who were not asked to exercise were far less likely to predict that they would regret not bringing water more than they would regret not bringing food.<sup>75</sup>

Another example of a projection-induced error is the *curse of knowledge*, where individuals permit their knowledge to interfere with their

---

<sup>72</sup> Russell W. Clement & Joachim Krueger, *Social Categorization Moderates Social Projection*, 38 J. EXPERIMENTAL SOC. PSYCHOL. 219, 219 (2002).

<sup>73</sup> Leaf Van Boven & George Loewenstein, *Cross-Situational Projection*, in THE SELF IN SOCIAL JUDGMENT 43, 49 (Mark D. Alicke et al. eds., 2005).

<sup>74</sup> *Id.*

<sup>75</sup> *Id.* at 51-52.

attempts to simulate others who are not only less knowledgeable, but whom they *know* are less knowledgeable, than they are.<sup>76</sup> One study, for example, asked well-informed businessmen to predict how individuals they knew were less informed would forecast corporate earnings. The study was designed to reward the mindreaders for quarantining their own superior knowledge; nevertheless, “the predictors failed to discount their own knowledge completely, so their predictions partly reflected their proprietary knowledge.”<sup>77</sup>

Such errors are troubling, because there is no reason to believe that jurors are any less prone to them than ordinary mindreaders—an issue discussed extensively below.<sup>78</sup> The existence of those errors, however, is still powerful evidence that jurors *do*, in fact, mindread through projection.<sup>79</sup> Indeed, functionalist theorizing cannot explain projection-based errors like empathy gaps and the curse of knowledge. Although a mindreader may well assume that a hiker in a hot emotional state will feel differently than a hiker in a cold emotional state, the attribution process is not supposed to be affected by which emotional state *the mindreader* is in—the same commonsense generalizations apply in both situations. Similarly, “[h]ow could an intuitive theory account for the curse of knowledge? Does everyone afflicted with the curse of knowledge accept as a general proposition that ‘other people believe whatever I believe’?”<sup>80</sup>

## 2. Jury Studies

Cognitive psychologists have not directly studied projection in the context of juror decision-making. A number of studies, however, provide strong circumstantial evidence that jurors sometimes use projection to mindread.

---

<sup>76</sup> Boaz Keysar et al., *States of Affairs and States of Mind: The Effects of Knowledge on Beliefs*, 64 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 283, 283 (1995).

<sup>77</sup> GOLDMAN, *supra* note 50, at 165-66.

<sup>78</sup> See *infra* Part V.A.

<sup>79</sup> See Goldman & Mason, *supra* note 3, at 285 (“These findings are congenial to simulation theory because one of the crucial underlying ideas of the theory is that mindreading involves a projection, imputation, or transference of one’s own mental state to the target.”).

<sup>80</sup> GOLDMAN, *supra* note 50, at 168. According to Van Boven and Loewenstein, the curse of knowledge “stems partly from the hindsight bias. That is, informed people’s biased predictions of the knowledge of uninformed individuals stems from informed people’s biased ‘postdictions’ of what their knowledge was before being informed.” Van Boven & Loewenstein, *supra* note 73, at 48. That is an explicitly projection-based explanation of the curse, and one that is supported by empirical evidence. See *id.* (noting that the thesis “that informed participants’ hindsight bias produced a corresponding curse of knowledge is supported by the close correspondence between self- and social-predictions”).

## a. Consent

Research into sexual intent attributions is particularly illuminating. Olsen-Fulero and Fulero have found, for example, that the most important variable minimizing the likelihood that a juror will blame a rape victim for being raped is “the degree to which a juror identifies with the victim . . . because they are similar or in some other way appealing . . . .”<sup>81</sup> They explained their finding by arguing that jurors who identify with victims blame them less for being raped because “they are more likely to be able to empathize with them and take their role so that the actor-observer difference is reduced . . . .”<sup>82</sup> In other words: (1) jurors who identify with the victim find it easier to put themselves in the victim’s position than those who do not; and (2) the easier it is for jurors to put themselves in the victim’s position, the more likely they are to understand and agree with the victim’s account of what happened—namely, that she did not consent. That explanation is consistent with the projection model, particularly its emphasis on perspective-taking and its assumption that identification is based on similarity between a juror and his mindreading target.<sup>83</sup>

Research by Lenton and her colleagues further supports the idea that jurors use projection to make mental-state determinations in rape cases. In their study, college students were divided into two groups, those who were interested in having casual sex and those weren’t. The two groups were then asked to read a scenario about a man and a woman on a first date and assess the character’s sexual intent. The researchers found that students in the “seeking casual sex” category imputed higher levels of sexual intent to the characters than the students in the “not seeking casual sex” category—a finding they specifically attributed to projection: “men’s and women’s judgments of sexual intent generally arise from one common projective process in which perceivers use their own standing on some issue to infer that of others.”<sup>84</sup>

To be sure, the study did not specifically address mindreading in a criminal case, thus providing only indirect support for the idea that *jurors* mindread through projection. The researchers made clear, though, that they believed their findings applied to jury decision-making: their interest in the relationship between projection and judgments of sexual intent stemmed

---

<sup>81</sup> Lynda Olsen-Fulero & Solomon M. Fulero, *Commonsense Rape Judgments: An Empathy-Complexity Theory of Rape Juror Story Making*, 3 PSYCHOL. PUB. POL’Y & L. 402, 409 (1997).

<sup>82</sup> *Id.*

<sup>83</sup> See *infra* notes 153-56, 218-32 and accompanying text.

<sup>84</sup> Alison P. Lenton et al., *We Want the Same Thing: Projection in Judgments of Sexual Intent*, 33 PERSONALITY & SOC. PSYCHOL. BULL. 975, 977 (2007).

from their recognition that “*mis*judgments may be related to sexual coercion, sexual harassment, and rape . . . .”<sup>85</sup>

b. Murder

Additional evidence for projection comes from a cross-cultural study of murder attributions by Morris and Peng. In the study, American and Chinese graduate students were asked to read newspaper reports about actual murders and evaluate whether the killings resulted from the killer’s disposition (character) or situation (external circumstances). They were then presented with counterfactual scenarios and asked to imagine whether the murders would still have occurred. As the researchers expected, culture played a significant role in the results. In the first task, “Chinese people represent[ed] behavior as situationally caused and Americans represent[ed] it as dispositionally caused.”<sup>86</sup> In the second task, “Chinese subjects simulated that the person would have taken a less bloody course of action in different situations, whereas Americans simulated that the person’s murderous disposition would have inexorably expressed itself, regardless of changes in the situation.”<sup>87</sup>

Despite their use of the term “simulation,” Morris and Peng did not specifically conclude that the graduate students determined the killers’ mental states through projection.<sup>88</sup> That is, however, a reasonable inference from their results. First, Morris and Peng rejected the idea that prototyping could explain the graduate students’ very different cultural attributions.<sup>89</sup> Second, other researchers *have* used projection to explain why situation/disposition mindreading attributions are so deeply influenced by culture:

If one perceives oneself as embedded within a larger context of which one is an interdependent part, it is likely that other objects or events will be perceived in a similar way. For example, a given event involving a particular actor will be perceived

---

<sup>85</sup> *Id.* at 975.

<sup>86</sup> Michael W. Morris & Kaiping Peng, *Culture and Cause: American and Chinese Attributions for Social and Physical Events*, 67 *J. PERSONALITY & SOC. PSYCHOL.* 949, 964 (1994).

<sup>87</sup> *Id.* at 964-65.

<sup>88</sup> It is also important to note that the disposition-situation distinction does not directly correlate with specific mental states. The study is still useful, however, because it clearly required the graduate students to understand the thought processes of the killers—did the killers “intend” to kill, or were they “compelled” to kill by the force of circumstance?

<sup>89</sup> Morris & Peng, *supra* note 86, at 965 (noting that “scripts would not lead to a general pattern of greater dispositionalism across everyday behavioral settings,” and that whereas “[a]tributitional patterns due to stereotypes would be restricted in scope to particular types of social actors or groups . . . . [C]ultural differences in attribution are far broader . . . .”).

as arising from the situational context of which this actor is an interdependent part, rather than as stemming solely from the attributes of the actor.<sup>90</sup>

That explanation makes sense. If individuals mindread by running simulations in which they apply their own “psychological mechanisms” to the target’s situation, an individual who sees his own actions as constrained by situational factors will be far more likely to attribute a situationally-influenced mental state to the target than an individual who believes that his actions are produced by his internal disposition.

### c. Mistake

Projection also likely explains the results of an experiment by Finkel and his colleagues that used a fictionalized version of the Bernhard Goetz murder trial to study the impact of mistake on self-defense claims. In the baseline scenario, a teenager stood over the defendant, pointed a gun at his head, and demanded five dollars; the defendant then pulled a gun out of his jacket and killed the youth—a classic self-defense situation, and one that led mock jurors to acquit 62.5% of the time. The researchers then presented the mock jurors with four variations on the baseline scenario in which increasingly unlikely mistakes led the defendant to kill: a *reasonable* mistake, in which the victim turned out to be carrying a lifelike toy gun; a *dubious* mistake, in which the victim was not carrying a gun but seemed to be reaching into his pocket for something; an *unreasonable* mistake, in which the victim was not carrying a gun and did not reach into his pocket; and finally a *delusional* mistake, in which the victim was sitting on the other side of the subway car doing a crossword puzzle and occasionally making eye contact with the defendant.<sup>91</sup> Not surprisingly, as the mistake became increasingly implausible, jurors became increasingly skeptical of the defendant’s claim that he honestly believed he was acting in self-defense: 79% acquitted the reasonable mistake; 25% acquitted the dubious mistake; 4% acquitted the unreasonable mistake; and no one acquitted the delusional mistake.<sup>92</sup>

The question, of course, is how the mock jurors determined whether the defendant honestly believed that he was about to be killed. Functionalist theorizing seems an unlikely candidate, for two reasons. First, the behavioral evidence in the scenarios was ambiguous: the killing itself was consistent with both an intentional killing and self-defense—which is

---

<sup>90</sup> Hazel Rose Markus & Shinobu Kitayama, *Culture and the Self: Implications for Cognition, Emotion, and Motivation*, 98 PSYCHOL. REV. 224, 246 (1991).

<sup>91</sup> Norman J. Finkel et al., *Lay Perspectives on Legal Conundrums: Impossible and Mistaken Act Cases*, 19 LAW & HUM. BEHAV. 593, 603-04 (1995).

<sup>92</sup> *Id.* at 605.

why the researchers chose it—and the non-unanimous verdicts for all but the delusional scenario indicate that there was no one-to-one correspondence between the circumstances that led to the killing and the defendant's mental state. (Even the classic self-defense baseline scenario led nearly 40% of the jurors to convict.) Second, jurors would have to possess a remarkably complex web of commonsense generalizations to distinguish so carefully between the scenarios—precisely the kind of “information rich” explanation that “attributes too much knowledge and sophistication to mindreaders.”

The results make complete sense, however, if jurors used projection to mindread the defendant. As Finkel has pointed out, a mistake defense requires jurors “to climb down and stand in the shoes and subjectivity of the defendants, in order to see what they see,”<sup>93</sup> because “the defendant's past experience, developmental history and acculturation, traumas and scars all may play a part in how he or she perceives events.”<sup>94</sup> Moreover, killing in self-defense seems like the kind of situation that a juror could convincingly simulate, even if (as is likely) he had never been in a similar situation himself. Finally, the mock jurors were not instructed to acquit only those mistakes that were *reasonable*,<sup>95</sup> a requirement that may push jurors away from projection and toward prototyping.<sup>96</sup> It thus makes sense to assume that the mock jurors determined whether the defendant honestly acted in self-defense by imagining themselves in the various situations, simulating their interaction with the victim, and determining whether *they* would have honestly believed they were about to be killed.

#### d. Capital Sentencing

Finally, Haney has specifically used projection to explain how jurors decide whether to sentence a defendant convicted of capital murder to death, a decision that requires them to assess the defendant's “personal culpability” for the crime.<sup>97</sup> Haney and his colleagues asked capital jurors to sentence a convicted murderer on the basis of extensive information concerning the nature of the crime and the murderer's life-history. In later interviews, a number of the jurors “gave accounts of having engaged in a

---

<sup>93</sup> FINKEL, *supra* note 28, at 229.

<sup>94</sup> *Id.* at 230.

<sup>95</sup> See Finkel et al., *supra* note 91, at 606 (“It is important to note that subjectivity's limits were not imposed by the law through judicial instructions, as no legal instructions were given in either experiment. These limits, it would seem, derive from ordinary common sense.”).

<sup>96</sup> See *infra* notes 168-75 and accompanying text.

<sup>97</sup> Craig Haney, *Commonsense Justice and Capital Punishment: Problematizing the “Will of the People,”* 3 PSYCHOL. PUB. POL'Y & L. 303, 330 n.109 (1997).

decision-making process in which they embraced a narrative version of the defendant's life, put themselves in his shoes, as it were, and came to an empathetic understanding of his social history from a largely subjective perspective."<sup>98</sup>

## V. PROTOTYPING

### A. PROTOTYPING DESCRIBED

First and foremost, substantive criminal law is concerned with defining the elements of specific crimes.<sup>99</sup> Some crimes require nothing more than conduct and a mental state, while others also require the presence of certain circumstances and/or the creation of certain harmful results.<sup>100</sup> Regardless of how they are defined, though, due process requires the prosecution to prove all of the elements of a crime "beyond a reasonable doubt."<sup>101</sup> Individual crimes are thus bundles of conditions that are "singly necessary and jointly sufficient": the failure to prove *any* element of a crime, no matter how seemingly insignificant, entitles the defendant to acquittal.<sup>102</sup>

A particular model of jury decision-making is implicit in this formal approach to criminal responsibility. Specifically, the approach assumes (1) that jurors actually understand crimes as bundles of "singly necessary and jointly sufficient" elements, and (2) that jurors will convict only when the prosecution proves all of those elements beyond a reasonable doubt. Empirical research, however, contradicts both assumptions. To begin with, jurors mentally represent crimes as *loosely-structured prototypes*, not as bundles of essential elements:

[T]he features contained in people's naive representations of crime categories do not operate as necessary conditions. Rather, these naive concepts appear to contain prototypes (or typical exemplars) of crime categories. Prototypes are summary representations of the characteristics of category members.<sup>103</sup>

---

<sup>98</sup> *Id.* at 329.

<sup>99</sup> See LAFAVE, *supra* note 1, § 1.2(a), at 7.

<sup>100</sup> *Id.* at 8-9.

<sup>101</sup> See, e.g., *Patterson v. New York*, 432 U.S. 197, 210 (1977) ("[T]he Due Process Clause requires the prosecution to prove beyond a reasonable doubt all of the elements included in the definition of the offense of which the defendant is charged.").

<sup>102</sup> See Vicki L. Smith, *When Prior Knowledge and Law Collide: Helping Jurors Use the Law*, 17 LAW & HUM. BEHAV. 507, 509 (1993).

<sup>103</sup> *Id.* at 509; see also Nancy Cantor et al., *A Prototype Analysis of Psychological Situations*, 14 COGNITIVE PSYCHOL. 45, 46-47 (1982) ("[K]nowledge about any given category is structured around, and represented in long-term memory as, a *prototype* which captures the meaning of the category . . . . [One form of the prototype could be] an abstract set of features commonly associated with members of a category, with each feature assigned a weight according to degree of association with the category." (citations omitted)).

Research indicates that jurors possess prototypes for a wide variety of crimes, including, *inter alia*, assault,<sup>104</sup> burglary,<sup>105</sup> kidnapping,<sup>106</sup> murder,<sup>107</sup> robbery,<sup>108</sup> stalking,<sup>109</sup> rape,<sup>110</sup> manslaughter,<sup>111</sup> euthanasia,<sup>112</sup> and infanticide.<sup>113</sup> These prototypes rarely correspond to the legal definition of a crime, an issue explored in detail below.<sup>114</sup> The most common juror prototype for assault, for example, is simply a physical attack that injures the victim,<sup>115</sup> a prototype that completely neglects the requirement that the victim reasonably fear bodily harm.<sup>116</sup>

Even more troubling, a number of studies have shown that jurors do not view the features associated with a particular prototype as “singly necessary and jointly sufficient”: if the characteristics of the defendant’s crime match the features of the prototype closely enough, jurors will convict even in the absence of one (or more) of the features.<sup>117</sup> Smith, for example, presented mock jurors with fourteen scenarios involving five different kinds of crimes—assault, burglary, kidnapping, murder, and robbery—and asked them to determine the defendant’s guilt.<sup>118</sup> Each

<sup>104</sup> Vicki L. Smith, *Prototypes in the Courtroom: Lay Representations of Legal Concepts*, 61 J. PERSONALITY & SOC. PSYCHOL. 857, 860 (1991).

<sup>105</sup> *Id.*

<sup>106</sup> *Id.*

<sup>107</sup> *Id.*

<sup>108</sup> *Id.*

<sup>109</sup> Susan Dennison & Donald M. Thomson, *Community Perceptions of Stalking: What Are the Fundamental Concerns?*, 7 PSYCHIATRY PSYCHOL. & L. 159, 166 (2000).

<sup>110</sup> Norman J. Finkel & Jennifer L. Groscup, *Crime Prototypes, Objective Versus Subjective Culpability, and a Commonsense Balance*, 21 LAW & HUM. BEHAV. 209, 224 (1997); Heather L. Littleton & Danny Axsom, *Rape and Seduction Scripts of University Students: Implications for Rape Attributions and Unacknowledged Rape*, 49 SEX ROLES 465, 470 (2003).

<sup>111</sup> Finkel & Groscup, *supra* note 110, at 219.

<sup>112</sup> *Id.* at 218.

<sup>113</sup> Norman J. Finkel et al., *Commonsense Judgments of Infanticide: Murder, Manslaughter, Madness, or Miscellaneous?*, 6 PSYCHOL. PUB. POL’Y & L. 1113, 1133 (2000).

<sup>114</sup> See *infra* notes 233-57 and accompanying text.

<sup>115</sup> See Smith, *supra* note 104, at 861 tbl.1.

<sup>116</sup> See, e.g., ILL. SUPREME COURT COMM. ON PATTERN JURY INSTRUCTIONS IN CRIMINAL CASES, ILLINOIS PATTERN JURY INSTRUCTIONS: IPI CRIMINAL § 11.02 (2d ed. 1981).

<sup>117</sup> See, e.g., Finkel & Groscup, *supra* note 110, at 212 (“Unlike the legal crime category of murder, for example, where a set of necessary and sufficient conditions determines category membership, ordinary citizens may not use a set of defining features in determining category membership, but a graded organization, where ‘prototypicality’ determines inclusion.”); Jennifer L. Skeem & Stephen L. Golding, *Describing Jurors’ Personal Conceptions of Insanity and Their Relationship to Case Judgments*, 7 PSYCHOL. PUB. POL’Y & L. 561, 569 (2001).

<sup>118</sup> Smith, *supra* note 104, at 860-61.

scenario omitted at least one essential element of the crime involved in the scenario, legally requiring the defendant's acquittal.<sup>119</sup> The results of the experiment were shocking:

The proportion of guilty votes was computed for each scenario and tested against zero. If a target feature is necessary, then subjects should vote not guilty when the feature is not present. To the extent that the conviction rate exceeds zero, subjects are not using the omitted feature as a necessary condition. Across the 14 critical scenarios, the conviction rates ranged from 71% to 100%, with a mode of 100%. All of the conviction rates significantly exceeded zero . . . , indicating that none of the frequently listed features was a necessary condition for guilt. It appears, then, that these attributes are associated with the crime categories but do not operate as defining features.<sup>120</sup>

Smith then conducted another experiment to test her proposed explanation of those results: namely, that because jurors "determine the guilt or innocence of a defendant by comparing the characteristics of the defendant's crime with the features of their prototype[,] [g]uilty verdicts would result if there was sufficient feature overlap [and] not-guilty verdicts would result if the correspondence was low."<sup>121</sup> That is exactly what the second experiment found—in each case "the typicality of the fact situation influenced subjects' verdict choices."<sup>122</sup>

The key question for our purposes, of course, is what role *mental-state* determinations play in the prototyping process. In terms of crime prototypes, there are two possible answers. The first and most troubling is that, for certain crimes, jurors simply make no attempt at all to mindread because their prototypes do not require the defendant to possess a particular mental state. Juror prototypes of stalking, for example, almost never mention the stalker intending to harm or cause fear in his target.<sup>123</sup>

Many crime prototypes, by contrast, do contain mens rea requirements. Most prototypes of murder, for example, include the defendant intending to kill the victim.<sup>124</sup> Indeed, such prototypes should be the rule, not the exception: "Generic representations include information regarding typical settings, typical characters, temporal sequence of events, causal relations, and the prototypic person's thoughts and feelings in these situations."<sup>125</sup>

---

<sup>119</sup> *Id.* at 861.

<sup>120</sup> *Id.* at 861-62.

<sup>121</sup> *Id.* at 859.

<sup>122</sup> *Id.* at 865.

<sup>123</sup> See *infra* notes 235-37 and accompanying text.

<sup>124</sup> See Smith, *supra* note 104, at 861 tbl.1; Richard L. Wiener et al., *The Psychology of Telling Murder Stories: Do We Think in Scripts, Exemplars, or Prototypes?*, 20 BEHAV. SCI. & L. 119, 131 tbl.3 (2002).

<sup>125</sup> Karniol, *supra* note 60, at 569; see also Cantor et al., *supra* note 103, at 65 ("[A] substantial part of the consensual descriptions of situations are psychological in nature; they

When a prototype does reference a particular mental state, the nature of prototyping suggests that jurors will treat the mental state as just another feature of the prototype, one of the many features that determine the overall typicality of the defendant's act.

That explanation, however, still implicates the fundamental mindreading question: how does a juror determine whether the defendant possessed the prototypical mental state? The evidence suggests two methods. To begin with, prototypes contain the "typical character" and the typical character's "thoughts and feelings" in the prototypic situation.<sup>126</sup> This suggests that if jurors conclude the defendant is the kind of person who typically commits a particular crime, they will also conclude that he possessed the mental state associated with the prototype. Indeed, that is exactly what research indicates. A study of infanticide prototypes, for example, found that defendants under age eighteen are much less likely to be convicted of first-degree murder—killing with premeditation—than defendants over the age of eighteen.<sup>127</sup> Similarly, another study found that black defendants are far more likely to be held criminally responsible for "prototypically black" crimes like soliciting, mugging, and auto theft than white defendants.<sup>128</sup>

Prototypes also include "the prototypic person's thoughts and feelings" in "typical settings."<sup>129</sup> This suggests that if jurors conclude that the defendant's actions are consistent with the prototypical situation, they will conclude that he possessed the prototypic mental state as well. Differently put, prototypes contain images of what particular mental states "look like"—images of how particular mental states are typically translated into action.<sup>130</sup> The infanticide research mentioned above, for example, found that jurors generally associate premeditation with particularly violent methods of killing, such as striking the baby with a blunt object, and the absence of premeditation with less violent methods, such as suffocation.<sup>131</sup>

---

include details about the most appropriate behavior for a situation, the feelings associated with 'being in a situation,' and the typical reactions and behavior of others in the situation.").

<sup>126</sup> Karniol, *supra* note 60, at 569; *see also* Cantor et al., *supra* note 103, at 59 n.3 ("Person images might include components such as physical appearance, personality characteristics, opinions, attitudes, beliefs, and typical behaviors, actions and interactions engaged in by the Person.").

<sup>127</sup> *See* Finkel et al., *supra* note 113, at 1130.

<sup>128</sup> *See* Christopher S. Jones & Martin F. Kaplan, *The Effects of Racially Stereotypical Crimes on Juror Decision-Making and Information-Processing Strategies*, 25 BASIC & APPLIED SOC. PSYCHOL. 1, 1 (2003).

<sup>129</sup> Karniol, *supra* note 60, at 569.

<sup>130</sup> For a fuller discussion of this issue, *see infra* notes 275-81 and accompanying text.

<sup>131</sup> *See* Finkel et al., *supra* note 113, at 1129.

The discussion thus far has focused specifically on crime prototypes. Jurors also appear to possess prototypes of certain *defenses*. Skeem and Golding, for example, designed a series of experiments to determine whether jurors assess a defendant's insanity defense by matching the defendant's characteristics to their prototype of the "typical" insane person. Here is how they described their hypothesis:

Just as jurors represent verdict categories in general criminal cases with crime prototypes, so too may jurors represent verdict categories in insanity cases with prototypes of insanity. Jurors may rely on their prototypes of insanity to construe the facts of the case and render a verdict on the basis of a prototype similarity matching process. A juror may make attributions about a defendant's cognitive and volitional impairments by comparing the defendant's characteristics to those of her prototype of the criminally insane defendant. The more closely the defendant's attributes match those of her prototype, the more likely she is to judge that defendant a member of the category "insane."<sup>132</sup>

The experiments strongly supported the hypothesis. In particular, Skeem and Golding concluded (1) that "jurors have complex, multifaceted prototypes of insanity that cannot be reduced to legal tests of insanity or to psychiatric diagnoses,"<sup>133</sup> and (2) that "prototypes of insanity are associated with case-relevant attitudes and strongly affect the way in which jurors construe case facts and render verdicts."<sup>134</sup>

Like crime prototypes, defense prototypes appear to take person-centered and situation-centered forms—and sometimes both. Juror prototypes of battered woman's syndrome, for example, are personological: the prototypical battered woman is young, fragile, guilt-ridden, confused, and depressed; rarely interacts with others; has children and is financially dependent on her partner; makes excuses for her partner's behavior; and accepts her abuse passively, without fighting back.<sup>135</sup> Juror prototypes of consent, by contrast, are primarily situational<sup>136</sup>: the prototypical woman who consents to sex flirted with her alleged rapist, went to his (or her) apartment, and engaged in consensual kissing or petting.<sup>137</sup> Finally, juror

---

<sup>132</sup> Skeem & Golding, *supra* note 117, at 569. "Category membership is based on a similarity matching process; as the number of features that the object shares with the category's prototype increases, so does the likelihood that the object is deemed a member of the category." *Id.* at 568.

<sup>133</sup> *Id.* at 604.

<sup>134</sup> *Id.*

<sup>135</sup> See Brenda L. Russell & Linda S. Melillo, *Attitudes Toward Battered Women Who Kill: Defendant Typicality and Judgments of Culpability*, 33 CRIM. JUST. & BEHAV. 219, 225-26 (2006).

<sup>136</sup> Consent prototypes also often contain personological aspects. For example, jurors strongly associate provocative dress with consent. See Littleton & Axsom, *supra* note 110, at 469 tbl.I.

<sup>137</sup> See *id.* at 472 tbl.III.

prototypes can be both personological and situational. Some jurors, for example, view a defendant as insane if he looks like the proverbial “wild beast,” while others base their insanity judgment on whether the defendant committed the crime in a suitably “bizarre” fashion.<sup>138</sup>

Finally, there is the issue of *how* jurors acquire the crime and defense prototypes they use to make mental-state determinations. With prototypes in general, there are two basic methods of acquisition: through experience and from exposure to “other sources,” such as the media.<sup>139</sup> Legal prototypes, however, almost certainly come primarily from the media, not from experience. As Valerie Hans has observed, “[b]ecause a relatively small proportion of the public has direct experience with the justice system, public knowledge and views of law and the legal system are largely dependent on media representations.”<sup>140</sup> That conclusion is supported by Holst and Pezdek’s study of robbery prototypes, which revealed that nearly 90% of the participants had learned what a “typical” robbery looked like from the media, with 50% discovering information from television shows, 23% from the news media, and 16% from movies.<sup>141</sup>

#### B. EVIDENCE OF PROTOTYPING

As a theory of mindreading, prototyping is based on two assumptions: (1) that jurors mentally represent crimes and defenses as prototypes, and (2) that jurors make at least some mental-state attributions on the basis of the correspondence between the evidence in the defendant’s case—situational and/or personological—and specific prototypes. I have already discussed much of the empirical research that supports the first assumption and have touched on the research that supports the second. In what follows, I will discuss other research that supports the idea that evidence-prototype matching often underlies jurors’ mindreading.

---

<sup>138</sup> See Caton F. Roberts et al., *Implicit Theories of Criminal Responsibility: Decision Making and the Insanity Defense*, 11 LAW & HUM. BEHAV. 207, 222-23 (1987).

<sup>139</sup> See, e.g., Jeffrey W. Sherman, *Development and Mental Representation of Stereotypes*, 70 J. PERSONALITY & SOC. PSYCHOL. 1126, 1127 (1996) (noting that the information underlying prototypes “may be acquired from firsthand personal experience with group members or through social learning from family, friends, and the media”).

<sup>140</sup> Valerie P. Hans, *Law and the Media: An Overview and Introduction*, 14 LAW & HUM. BEHAV. 399, 399 (1990); see Norman J. Finkel, *Achilles Fuming, Odysseus Stewing, and Hamlet Brooding: On the Story of the Murder/Manslaughter Distinction*, 74 NEB. L. REV. 742, 777 (1995); see also Smith, *supra* note 102, at 508 (noting that most jurors “have had some exposure to the law through informal sources, such as movies, television, newspapers, and word of mouth. From these sources, jurors may have abstracted representations of many legal concepts that they can access for decision making . . .”).

<sup>141</sup> Valerie Fisher Holst & Kathy Pezdek, *Scripts for Typical Crimes and Their Effects on Memory for Eyewitness Testimony*, 6 APPLIED COGNITIVE PSYCHOL. 573, 580 (1992).

### 1. Crimes

In one revealing study, Jones and Kaplan examined whether jurors would be more likely to hold defendants responsible for race-congruent crimes than race-incongruent crimes.<sup>142</sup> The study extended previous work indicating that the race of the defendant is one of the characteristic features of most crime prototypes.<sup>143</sup> The results were dramatic: "When charged with a racially congruent crime, the defendant is perceived to be more responsible for that crime and the behaviors tend to be attributed more to internal rather than to external factors."<sup>144</sup>

The infanticide study mentioned earlier also strongly supports evidence/prototype matching. The study was designed to elicit jurors' prototypes of infanticide and examine how those prototypes influenced their mental-state attributions.<sup>145</sup> To do so, the researchers presented college students with nine different versions of a fictional criminal case in which a woman is charged with first-degree murder for killing her newborn child.<sup>146</sup> The defendant claims that the killing was caused by her depression, negating premeditation and requiring the jury to find her either guilty of voluntary manslaughter or guilty but mentally ill. The nine versions differed in terms of four main variables: the extent of the defendant's depression (simple depression, postpartum depression, depression with psychosis); the time between birth and the murder (one day, one week, three months); the age of the woman (fourteen, nineteen, twenty-six); and the manner of death (abandonment, suffocation, blunt-force trauma). The jurors were asked to read each version of the case and reach the appropriate verdict.<sup>147</sup>

The results of the study revealed that the prototypic infanticide is a situation in which a young mother suffering from depression suffocates her baby almost as soon as it is born, and that infanticide verdicts are determined by the degree to which the facts of the defendant's case match that prototype. As the age of the defendant increased, so did attributions of

---

<sup>142</sup> Jones & Kaplan, *supra* note 128.

<sup>143</sup> See, e.g., Michael Sunnafrank & Norman E. Fontes, *General and Crime Related Racial Stereotypes and Influence on Juridic Decisions*, 17 CORNELL J. SOC. REL. 1, 10 (1983). They found that jurors viewed crimes like soliciting, assault, mugging, auto theft, and assault on a police officer as prototypically black, and viewed crimes like embezzlement, child molestation, counterfeiting, and fraud as prototypically white. *Id.*

<sup>144</sup> Jones & Kaplan, *supra* note 128, at 9. Interestingly, jurors who convicted defendants of crimes that matched their person-prototypes were also more confident in the defendant's guilt and punished the defendant more severely than jurors who convicted defendants of non-matching crimes. *Id.*

<sup>145</sup> Finkel et al., *supra* note 113, at 1121.

<sup>146</sup> *Id.* at 1122.

<sup>147</sup> *Id.* at 1122-23.

premeditation. Fourteen-year-old defendants were substantially more likely to be convicted of voluntary manslaughter or found guilty but mentally ill than nineteen- or twenty-six-year-old defendants.<sup>148</sup> Premeditation was also positively correlated with time—the greater the gap between the baby’s birth and death, the more willing jurors were to convict the defendant of first-degree murder.<sup>149</sup> Finally, the manner of the killing significantly affected premeditation attributions—abandoning the baby in a dumpster or beating it to death led to far more first-degree murder convictions than suffocating it.<sup>150</sup>

## 2. Defenses

Studies of defenses also reveal prototyping. Russell and Melillo, for example, used pre-existing research into prototypes of battered women—summarized above—to determine whether prototypicality affects mental-state attributions and verdict choices in cases involving battered women’s self-defense claims. The results were unequivocal:

As expected, participants in atypical/active conditions were more likely to assign judgments of second-degree murder, whereas those in typical/passive conditions most often rendered not-guilty verdicts. Atypical defendants were perceived to have more options available, and respondents were less likely to believe the defendant’s version of events. In contrast, respondents rated typical defendants as more likely to meet the requirements of self-defense.<sup>151</sup>

The researchers thus concluded that “judgments of culpability are indeed influenced by the extent to which the defendant fits or does not fit the typology of a battered woman.”<sup>152</sup>

## VI. PERCEIVED SIMILARITY AS THE MEDIATING MECHANISM

### A. THE MECHANISM

Although both are irreconcilable with functionalist theorizing, projection and prototyping could not be more different. A juror who mindreads through projection tries to emulate what the defendant saw, thought, and felt when he committed the crime; he literally tries to *become* the defendant. A juror who mindreads through prototyping, by contrast, has little interest in delving into the defendant’s subjective experience; instead, he determines the defendant’s mental state by appealing to his own existing

---

<sup>148</sup> *Id.* at 1126.

<sup>149</sup> *Id.* at 1124-25.

<sup>150</sup> *Id.* at 1129-30.

<sup>151</sup> Russell & Melillo, *supra* note 135, at 234.

<sup>152</sup> *Id.* at 235.

preconceptions of what the “typical” crime, defense, or defendant looks like.

As we will see, there is reason to believe that neither projection nor prototyping is particularly likely to result in accurate mental-state determinations. Before we reach that issue, however, we must first answer a critical question: what determines which mindreading technique a juror will use in a particular case?

The answer, according to Daniel Ames, is *perceived similarity*—the degree to which the juror perceives the defendant to be similar to himself.

I suggest that perceptions of similarity—an idiosyncratic and subjective sense that one is similar to a target group—moderate both projection and stereotyping in prevalence estimates of novel attributes. This yields what could be called a *similarity contingency model* of social inference . . . : When perceivers assume higher levels of general similarity to a target group, they engage in higher levels of projection on specific attributes, introspecting about their own attitudes and qualities and ascribing them to the target . . . . When perceivers assume lower levels of general similarity to a target, they engage in higher levels of stereotyping, turning to implicit beliefs about what a particular group is like . . . .<sup>153</sup>

General mindreading research supports that explanation. A series of studies conducted by Ames himself, for example, found that perceived similarity mediates between projection and prototyping “in domains ranging from romantic interactions to face-to-face business negotiations.”<sup>154</sup> Similarly, Clement and Krueger have demonstrated, through two experiments that manipulated participants’ in-group and out-group categorizations, that “self-referent knowledge serves as a readily accessible anchor for in-group estimates but . . . is suspended for out-group estimates.”<sup>155</sup> Finally, Lenton and her colleagues have established “that projection of sexual intent is most likely to occur when the perceiver believes the target to be similar to himself or herself.”<sup>156</sup>

None of these findings should come as a surprise. If the goal of projection is to emulate what the defendant saw, thought, and felt at the

---

<sup>153</sup> Daniel R. Ames, *Strategies for Social Inference: A Similarity Contingency Model of Projection and Stereotyping in Attribute Prevalence Estimates*, 87 J. PERSONALITY & SOC. PSYCHOL. 573, 574 (2004); cf. Clement & Krueger, *supra* note 72, at 219 (“There is a striking exception [to mindreaders’ general tendency to project]. The boundaries of social categories are veritable firewalls against the spread of projection.”); Karniol, *supra* note 60, at 567 (suggesting that “in those domains in which self is perceived as distinct rather than as similar to most others, self is not used to make consensus estimates”).

<sup>154</sup> Daniel R. Ames, *Inside the Mind Reader’s Tool Kit: Projection and Stereotyping in Mental State Inference*, 87 J. PERSONALITY & SOC. PSYCHOL. 340, 350 (2004).

<sup>155</sup> Clement & Krueger, *supra* note 72, at 228.

<sup>156</sup> Lenton et al., *supra* note 84, at 984; see also Dennis Krebs, *Empathy and Altruism*, 32 J. PERSONALITY & SOC. PSYCHOL. 1134, 1143 (1975) (concluding “that perception of similarity increases the disposition to imagine how one would feel in another’s place”).

time of the crime, a juror must believe *ex ante* that such emulation is possible—that the defendant generally sees, thinks, and feels like he does. That belief, in turn, is likely to exist only if the juror perceives himself to be similar to the defendant; in the absence of perceived similarity, the juror will have no reason to assume that he can use his own perceptions, thoughts, and feelings as the starting point of the mindreading process. He will thus turn instead to prototyping—his pre-existing understanding of how people like the defendant “typically” act and what a crime like the one the defendant is accused of committing “normally” looks like.<sup>157</sup>

It is important to emphasize here that the mediating factor is *perceived* similarity, not *actual* similarity. As Ames notes, “perceived general similarity is not expected to track closely, or even at all, with measures of actual similarity.”<sup>158</sup> Indeed, empirical work suggests that it doesn’t.<sup>159</sup> It is thus difficult to predict what factors will lead a particular juror to see himself as similar to, or different from, a defendant. We would expect sociological differences to be critical: whites perceiving themselves as different from blacks, men perceiving themselves to be different from women, etc. It is clear, though, that such differences do not *always* lead jurors to prototype. A study conducted by Van Boven and Loewenstein, for example, found that men were willing to simulate a woman’s experience of childbirth and women were willing to simulate a man’s experience of testicular cancer.<sup>160</sup> Similarly, although white mock jurors used prototyping to assess the culpability of black defendants in Jones and Kaplan’s study, they also used prototyping to assess the culpability of white defendants.<sup>161</sup>

---

<sup>157</sup> See, e.g., Epley et al., *supra* note 68, at 337 (“[I]t is reasonable to assume that people engage in such anchoring and adjustment only when it is clear that one’s self is a reasonable starting place in estimating others’ perspectives. In situations where one’s own perspective is clearly irrelevant, quite different processes may be involved . . . . Instead, each may generate an assessment of the other person’s views based on some stored representation (e.g., a prototype) of their ideologies and values.”); Robert M. Gordon, *Folk Psychology as Simulation*, in *FOLK PSYCHOLOGY: THE THEORY OF MIND DEBATE* 60, 65-66 (Martin Davies & Tony Stone eds., 1995) (“Within a close-knit community, where people have a vast common fund of ‘facts’ as well as shared norms and values, only a minimum of pretending would be called for . . . . A person transplanted into an alien culture might have to do a great deal of pretending to explain and predict the behavior of those around him. Indeed, one might eventually learn to *begin* all attempts at explanation and prediction with a stereotypic set of adjustments . . . .”).

<sup>158</sup> Ames, *supra* note 153, at 575.

<sup>159</sup> See *id.* at 582.

<sup>160</sup> Van Boven & Loewenstein, *supra* note 73, at 46.

<sup>161</sup> Jones & Kaplan, *supra* note 128, at 5. It is possible, of course, that the white mock jurors were primed to use prototyping by the presence of black defendants and would have used projection instead had the study only involved white defendants. The study did not address that question.

There may also be a motivational component to perceptions of similarity. Studies have shown that individuals differ in their need to feel unique.<sup>162</sup> High-uniqueness jurors should be less likely to see themselves as similar to the defendant, and thus more likely to prototype instead of project, than low-uniqueness jurors. Indeed, studies by Ames and Iyengar have found exactly that.<sup>163</sup> Moreover, because all individuals have at least some need to feel unique,<sup>164</sup> jurors may be generally motivated to see themselves as different from criminal defendants, particularly those who are charged with heinous or disgusting crimes. If so, the very nature of criminal trials may predispose jurors to prototype instead of project. That might explain why jury decision-making studies have found far more evidence of prototyping than evidence of projection.<sup>165</sup>

#### B. NATURE OF THE CRIME OR DEFENSE

Although perceived similarity is clearly the most important determinant of whether a juror mindreads through projection or prototyping, that choice may also be affected by the nature of the specific mental-state determination that the juror is asked to make. There may be an elective affinity, in other words, between particular kinds of crimes and defenses and particular mindreading techniques. Specifically, it seems reasonable to suggest that the more subjective the mens rea, the more likely a juror will be to project.

Consider, for example, the subjective version of entrapment, which asks whether the defendant was predisposed to commit the induced offense or was simply an “unwary innocent”<sup>166</sup> whose will was overborne by the nature of the inducement.<sup>167</sup> It is unlikely that jurors possess prototypes of “typical” entrapment situations, given that few (if any) jurors will have any personal experience with entrapment and the defense is rarely the subject of movies, television, or news reports. It thus seems more plausible to assume

---

<sup>162</sup> See, e.g., C.R. Snyder & Howard L. Fromkin, *Abnormality as a Positive Characteristic: The Development and Validation of a Scale Measuring Need for Uniqueness*, 86 J. ABNORMAL PSYCHOL. 518, 525 (1977).

<sup>163</sup> Daniel R. Ames & Sheena S. Iyengar, *Appraising the Unusual: Framing Effects and Moderators of Uniqueness-Seeking and Social Projection*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 271, 280 (2005).

<sup>164</sup> Snyder & Fromkin, *supra* note 162, at 525.

<sup>165</sup> To be sure, cognitive psychologists who study jury decision-making may simply be more interested in prototyping. Ames notes, though, that “[a]lthough projection has received lavish scholarly attention as a source of mindreading, stereotyping has been almost entirely ignored.” Ames, *supra* note 54, at 163.

<sup>166</sup> *Sherman v. United States*, 356 U.S. 369, 372 (1958).

<sup>167</sup> See Eugene Borgida & Roger Park, *The Entrapment Defense: Juror Comprehension and Decision Making*, 12 LAW & HUM. BEHAV. 19, 21 (1988).

that a juror would determine the defendant's mental state by simulating the defendant's entrapment situation and determining whether he (the juror) would have been predisposed to commit the crime.

If a subjective mens rea promotes projection, an objective mens rea—a mens rea that includes a reasonableness requirement—may promote prototyping. A defendant is entitled to the defense of duress, for example, only if a reasonable person in his situation would have been unable to resist the external pressure.<sup>168</sup> The very nature of the defense, in other words, communicates to jurors the need to consider *more* than just the defendant's (possibly idiosyncratic) psychological responses. That message may well lead jurors to prototype instead of project: given that most legal prototypes are disseminated by the media, jurors are likely to see them as far more representative of how people normally react in certain circumstances than their own (possibly idiosyncratic) reactions.

To be sure, this interpretation of reasonableness requirements is speculative. It is possible that jurors use projection to mindread even when a crime's mens rea requirement is objective, assuming that their (imagined) reaction to the defendant's situation would be representative of people generally. One study of provocation has found that "individuals integrate, to some extent, their own distinctive identity and characteristics when constructing ordinariness."<sup>169</sup> And, as noted earlier, "[w]hen people predict the thoughts, feelings, or behaviors of others, they tend to assume that these others think, feel, and behave as they themselves do."<sup>170</sup>

Notice, though, that this form of projection differs from the kind of projection discussed earlier, in which a juror uses simulation to determine the mental state he would have had in the defendant's situation *if he were the defendant*, not the mental state *he would have had* in the defendant's situation. Only such non-adjusted projection—projection in which the juror makes no attempt to emulate the thought processes of the defendant—would be useful for reasonableness determinations. Do jurors engage in both forms of projection? Additional research is necessary, but it is worth noting that at least one simulation theorist has argued that non-adjusted projection is "the default mode of simulation."<sup>171</sup> Moreover, as we will see in the next Part, jurors, and individuals generally, seem to have great difficulty making the adjustments that are necessary to ensure accurate projection.

---

<sup>168</sup> See JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 23.02, at 326-28 (4th ed. 2006).

<sup>169</sup> Peter Papanthasiou & Patricia Easta, *The 'Ordinary Person' in Provocation Law: Is the 'Objective' Standard Objective?*, 11 CURRENT ISSUES IN CRIM. JUST. 53, 62 (1999).

<sup>170</sup> Clement & Krueger, *supra* note 72, at 219.

<sup>171</sup> Gordon, *supra* note 70, at 13.

Finally, there is the question of how jurors will deal with what Paul Robinson has called a “subjectivized standard of reasonableness”<sup>172</sup>—a standard that requires the defendant’s mental state to be reasonable, but to some extent determines reasonableness from the defendant’s perspective.<sup>173</sup> The Model Penal Code’s “extreme emotional disturbance” defense is an example: although there must be a “reasonable explanation” for the disturbance, “[t]he reasonableness of such explanation or excuse shall be determined from the viewpoint of a person in the actor’s situation under the circumstances as he believes them to be.”<sup>174</sup> Another example is the approach taken to self-defense by cases such as *State v. Wanrow*, in which the Supreme Court of Washington held that a defendant is entitled to use the amount of force that “a reasonable person in the same situation . . . seeing what (s)he sees and knowing what (s)he knows, then would believe to be necessary.”<sup>175</sup> It is difficult to predict what mindreading technique jurors would use when faced with this kind of subjectivized reasonableness standard. Its overall objectivity would seem to encourage prototyping, but its subjective component would seem to encourage projection.

## VII. THE ERRORS OF PROJECTION AND PROTOTYPING

The discussion thus far has focused on how jurors mindread through projection and prototyping. We now turn to an equally critical question: are projection and prototyping likely to result in *accurate* mental-state determinations? Unfortunately, there is reason to be skeptical.

### A. THE ERRORS OF PROJECTION

We have seen that, as a mindreading technique, projection involves an attempt by the juror to imagine not how he would have felt in the defendant’s situation, but how he would have felt in the defendant’s situation *if he were the defendant*. The two imaginary processes are analytically different—and empirical research has shown that they are *phenomenologically* different, as well: “imagining how another feels and

---

<sup>172</sup> See PAUL H. ROBINSON, CRIMINAL LAW 538-39 (1997).

<sup>173</sup> See Kevin Jon Heller, *Beyond the Reasonable Man? A Sympathetic but Critical Assessment of the Use of Subjective Standards of Reasonableness in Self-Defense and Provocation Cases*, 26 AM. J. CRIM. L. 1, 5 (1998).

<sup>174</sup> MODEL PENAL CODE & COMMENTARIES § 210.3(b) (1985). The MPC’s definitions of recklessness and negligence also contain subjectivizing language. See *id.* § 2.02(c) (recklessness); *id.* § 2.02(d) (negligence).

<sup>175</sup> 559 P.2d 548, 557 (Wash. 1977).

imagining how you would feel . . . produce different emotional consequences."<sup>176</sup>

There is no guarantee, however, that a juror's attempt "to make his own mind 'emulate' the mental sequence the target will go through" will succeed. On the contrary, *accurate* projection has two necessary preconditions:

In the mindreading case, process-driven simulation can succeed in producing a final state that is identical or isomorphic to that of the target as long as (1) the process or mechanism driving the simulation is identical, or relatively similar, to the process or mechanism that drives the target and (2) the initial states of the simulating system (the attributor) are the same as, or relevantly similar to, those of the target.<sup>177</sup>

For projection to be accurate, in other words, a juror must not only *perceive* himself to be similar to the defendant, he must also *actually be similar* to the defendant when he engages in the act of mindreading. The perceptions and reactions he uses to simulate the defendant's situation must match the perception and reactions of the defendant in the actual situation.<sup>178</sup>

Is the requisite degree of similarity likely to exist between jurors and defendants? To answer that question, we must first distinguish between two kinds of similarity. We might call the first *natural similarity*: a situation in which the juror and the defendant are psychologically similar simply by virtue of their shared sociological characteristics. In such a situation, a juror will be able to accurately mindread the defendant simply by creating a simulation that accurately reproduces the external

<sup>176</sup> C. Daniel Batson et al., *Perspective Taking: Imagining How Another Feels Versus Imagining How You Would Feel*, 23 J. PERSONALITY & SOC. PSYCHOL. 751, 756 (1997).

<sup>177</sup> GOLDMAN, *supra* note 50, at 32; *see also* Paul L. Harris, *From Simulation to Folk Psychology: The Case for Development*, in FOLK PSYCHOLOGY: THE THEORY OF MIND DEBATE, *supra* note 157, at 207, 217 ("[I]t is necessary for the simulator to feed in pretend inputs that match in the relevant particulars the situation facing the agent whose actions are to be predicted or explained. Predictive errors will occur if inappropriate pretend inputs are fed in."); Josef Perner, *Simulation as Explicitation of Predication-Implicit Knowledge About the Mind: Arguments for a Simulation-Theory Mix*, in THEORIES OF THEORIES OF MIND 90, 94 (Peter Carruthers & Peter K. Smith eds., 1996) ("[I]t is imperative that the sequence of predication-implicitly represented mental states are real sequences executed off-line and which are relevantly similar to the sequence of mental states in the simulated person. This means they function as an analogue model.").

<sup>178</sup> *See, e.g.*, Alvin I. Goldman, *Interpretation Psychologized*, in FOLK PSYCHOLOGY: THE THEORY OF MIND DEBATE, *supra* note 157, at 74, 89 ("[I]t may well be considered remarkable that we are able to predict human behavior as well as we do. Is this impressive success fully accounted for by the simulation theory? Only, I think, with an added assumption, viz., that the other people, whose behavior we predict, are psychologically very similar to ourselves."); Perner, *supra* note 177, at 93 ("The critical condition for [simulation] having any chance of success is that it exploits (wittingly or not) the fact that one's own psychological make-up is relevantly similar to the make-up of the person simulated.").

circumstances that led to the defendant's crime; no adjustments to "the process or mechanism driving the simulation" will be required.

As we will see, such natural similarity is likely the exception, not the rule. In most cases, because jurors and the defendant will be very sociologically different, they will be very psychologically different, as well. To mindread accurately, therefore, jurors will normally have to *adjust the simulation process* to eliminate—or at least minimize—those psychological differences, thereby creating what we might call *constructed similarity*.<sup>179</sup>

### 1. Natural Similarity

A complete account of the sociological differences between jurors and defendants is beyond the scope of this Article. Significant differences clearly exist, however, as a result of problems with jury-selection procedures<sup>180</sup> and the sociologically-specific use of peremptory challenges.<sup>181</sup> Jurors are disproportionately middle-aged, middle-class, white,<sup>182</sup> and female.<sup>183</sup> Criminal defendants, by contrast, are disproportionately "young, poor, of color,"<sup>184</sup> and male.<sup>185</sup> In seventeen trials involving black male defendants held in Champaign County, Illinois,

---

<sup>179</sup> See, e.g., Gordon, *supra* note 157, at 63 ("As in the case of hypothetical self-prediction, the methodology essentially involves deciding what to do; but, extended to people of 'minds' different from one's own, this not the same as deciding what I myself would do. One tries to make adjustments for relevant differences." (emphasis omitted)); Van Boven & Loewenstein, *supra* note 73, at 45 ("[P]erspective taking requires the realization and anticipation that because other people are in a different situation than the self—whether that difference concerns expertise, bodily drives, or emotional reactions—they will think and feel differently than the self.").

<sup>180</sup> See, e.g., Darryl K. Brown, *The Means and Ends of Representative Juries*, 1 VA. J. SOC. POL'Y & L. 445, 451-61 (1994).

<sup>181</sup> See, e.g., Shaun L. Gabbidon et al., *Race-Based Peremptory Challenges: An Empirical Analysis of Litigation from the U.S. Court of Appeals, 2002-2006*, 33 AM. J. CRIM. JUST. 59 (2008).

<sup>182</sup> See Brown, *supra* note 180, at 446 (noting that "racial and ethnic minorities, as well as the young, old, and poor, are constantly underrepresented in most federal and state court jury pools and wheels").

<sup>183</sup> See, e.g., LEAGUE OF WOMEN VOTERS OF CHAMPAIGN COUNTY & UNIV. OF ILL. COLLEGE OF LAW, CHAMPAIGN COUNTY COURTWATCHING PROJECT HIGHLIGHTS: THIRD ANNUAL REPORT 2006-2007 (2007) (noting that in Champaign County, women were significantly overrepresented on federal and state juries).

<sup>184</sup> Eva S. Nilsen, *Decency, Dignity, and Desert: Restoring Ideals of Humane Punishment to Constitutional Discourse*, 41 U.C. DAVIS L. REV. 111, 172 (2007); see also CAROLINE WOLF HARLOW, U.S. DEP'T OF JUSTICE, DEFENSE COUNSEL IN CRIMINAL CASES 1 (2000) (noting that "[a]bout 80% of all criminal defendants are poor").

<sup>185</sup> See, e.g., LEAGUE OF WOMEN VOTERS OF CHAMPAIGN COUNTY & UNIV. OF ILL. COLLEGE OF LAW, CHAMPAIGN COUNTY COURTWATCHING PROGRAM HIGHLIGHTS: SECOND ANNUAL REPORT 2005-2006 (2006) (noting that in 2005, 93% of defendants in Champaign County federal and state courts were male).

for example, only 4 out of 252 jurors and alternates were black men and only 10 out of 252 were black women.<sup>186</sup>

These sociological differences would not matter if there was no causal relationship between an individual's psychology and his race, gender, nationality, and so on. But that is clearly not the case: the empirical literature is legion with examples of what Kim Lane Scheppelle has called the "perceptual fault lines that occur at the boundaries between social groups, between whites and people of color, between the privileged and the poor, between men and women"<sup>187</sup>—fault lines that have a dramatic impact on how jurors assess defendants' culpability. White jurors, for example, are more likely to convict black defendants than white defendants, are more likely to convict a defendant whose victim was white than a defendant whose victim was a minority, and tend to punish black defendants more severely.<sup>188</sup> Wealthy jurors are more likely to convict than poor jurors, and jurors are generally more likely to convict a defendant from a different socioeconomic class than a defendant from their own class, a bias that magnifies as the disparity between the juror's and the defendant's socioeconomic class increases.<sup>189</sup> And female jurors are more conviction-prone than male jurors in rape and incest cases, but are generally more lenient toward defendants in cases that don't involve sex.<sup>190</sup>

At least some of these differences likely result from the effect of sociological differences on mindreading. We saw earlier that Chinese graduate students were far more willing than their American counterparts to attribute a defendant's decision to kill to situational pressures.<sup>191</sup> A number

---

<sup>186</sup> LEAGUE OF WOMEN VOTERS OF CHAMPAIGN COUNTY & UNIV. OF ILL. COLLEGE OF LAW, REPORT ON THE 2004-2005 PILOT PHASE OF THE CHAMPAIGN COUNTY COURTWATCHING PROJECT 2 (2005).

<sup>187</sup> Kim Lane Scheppelle, *Foreword: Telling Stories*, 87 MICH. L. REV. 2073, 2083 (1989).

<sup>188</sup> Heller, *supra* note 173, at 50.

<sup>189</sup> *Id.* at 50-51.

<sup>190</sup> *Id.* at 50.

<sup>191</sup> See *supra* notes 87-88 and accompanying text. This situation-disposition split applies to Western and Eastern cultures in general. See, e.g., Fiona Lee et al., *Explaining Real-Life Events: How Culture and Domain Shape Attributions*, 22 J. PERSONALITY & SOC. PSYCHOL. 732, 733 (1996) ("Numerous studies have found that people from Western cultures tend to make more dispositional attributions, whereas people from Eastern cultures tend to make more situational attributions."). The situation-disposition split has even been found *within* Western cultures. See, e.g., Angeline Lillard, *Developing a Theory of Mind: The CIAO Approach*, 18 CURRENT DIRECTIONS IN PSYCHOL. SCI. 57, 59 (1999) ("The urban children used psychological explanations frequently and early; about 60% of their explanations for others' good and bad behaviors were psychological even at age 7 . . . . In contrast, the rural children averaged only 20% psychological explanations, and instead used mostly situational explanations.").

of cross-cultural mindreading studies have reached similar conclusions. Hamilton and Sanders, for example, found that Americans were more likely than Japanese to view a defendant's criminal act as intentional instead of as negligent.<sup>192</sup> Similarly, a study by Levinson found that "Chinese consistently made higher state of mind judgments than Americans, including judgments of extreme recklessness, intent to kill, purpose to kill, knowledge, and deliberation."<sup>193</sup>

It is not difficult to see how these psychological differences could undermine accurate projection. Consider, for example, a case in which a white juror is faced with an Asian defendant who is arguing subjective entrapment, a defense that depends on whether the defendant's criminal act is attributable to his predisposition to commit the crime or to the irresistible nature of the government agent's inducement.<sup>194</sup> As we have seen, projection research indicates that because whites are more likely than Asians to view *their own* actions as dispositionally caused, whites are generally more willing than Asians to see the actions of *others* as dispositionally caused. The white juror who uses projection to mindread the Asian defendant is thus likely to make an inaccurate mental-state attribution: whereas the white juror will likely view the defendant's act as dispositionally caused, the Asian defendant will likely view his act as caused by the government agent's inducement—he will not see himself as having been subjectively predisposed to commit the crime.

This does not mean that the white juror is doomed to misread the Asian defendant's mental state. It simply means that there is no *natural* similarity between the juror and the defendant that can guarantee accurate projection. *Constructed* similarity may still be possible: the white juror may be able to adjust his simulation of the defendant's experience to eliminate the perceptual gap that exists between them. To the possibility of such adjustment we now turn.

## 2. Constructed Similarity

Constructed similarity has two prerequisites. First, the juror must prevent his own idiosyncratic desires and beliefs and knowledge from influencing his simulation of the defendant's experience, a process that Goldman calls "quarantining."<sup>195</sup> If he does not quarantine his desires and beliefs and knowledge, the simulation routine will reveal what mental state

---

<sup>192</sup> See V. LEE HAMILTON & JOSEPH SANDERS, EVERYDAY JUSTICE 109 (1992).

<sup>193</sup> Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1, 21 (2005).

<sup>194</sup> See *supra* notes 166-67 and accompanying text.

<sup>195</sup> GOLDMAN, *supra* note 50, at 29.

he would have had in the defendant's situation, not what mental state the defendant actually had. Second, as noted above, the juror must adjust his simulation routine to take into account the defendant's psychological differences. If he does not make the necessary adjustments, the simulation will once again reveal his mental state, not the defendant's.<sup>196</sup> "Pretend inputs," in short, "can be inaccurate in two ways, through either an excess or a deficiency of inputs."<sup>197</sup>

#### a. Excess of Inputs

The fundamental problem with projection, at least in terms of constructed similarity, is that individuals *simply don't quarantine effectively*. The goal of quarantining is to allow the juror to recognize that the mindreading target may have very different thoughts, feelings, and knowledge than he does. It is this ideational space that makes it possible for the mindreader to subsequently adjust his simulation routine to take those differences into account.<sup>198</sup> Unfortunately, as noted earlier, mindreaders rarely acknowledge psychological difference; on the contrary, they assume that others "think, feel, and behave as they themselves do" even "when they are asked not to" and "regardless of information they have about other individuals."<sup>199</sup>

Even worse, research indicates that this stubborn egocentrism is at its apex when the target's behavior is *ambiguous*—the normal situation in a criminal case that focuses on the defendant's mental state.<sup>200</sup> Lenton and her colleagues, for example, have found that both men and women are particularly likely to project their own sexual intent onto others when others' sexual intent is uncertain.<sup>201</sup> More generally, Green and Sedikides have shown that "the influence of self-schemas is pervasive when the target's personality characteristics are ambiguous."<sup>202</sup> Neither finding is surprising: "[i]f one has no direct knowledge of what another . . . does or does not know, and little or no knowledge that would provide the basis for

---

<sup>196</sup> *Id.* at 29.

<sup>197</sup> *Id.* at 172.

<sup>198</sup> *Cf. id.* at 165 ("[W]e can profitably understand projection in terms of a quarantine-violating simulation process in which the quarantine violation strongly affects, or contaminates, the resulting attribution . . . . This results in an attribution that is inappropriately influenced by the attributor's own current states (genuine, nonpretend states).").

<sup>199</sup> Clement & Krueger, *supra* note 72, at 219.

<sup>200</sup> See *supra* notes 59-63 and accompanying text.

<sup>201</sup> See Lenton et al., *supra* note 84, at 984.

<sup>202</sup> Jeffrey D. Green & Constantine Sedikides, *When Do Self-Schemas Shape Social Projection?: The Role of Descriptive Ambiguity*, 25 *MOTIVATION & EMOTION* 67, 80 (2001).

making inferences in this regard, the only thing left to do is to use one's own knowledge as a default assumption as to what the other knows."<sup>203</sup>

The quarantining failure that underlies egocentrism leads to a number of important mindreading errors. Empathy gaps, mentioned earlier, are one example:

A growing body of evidence indicates that people in one emotional situation project their current preferences and behaviors onto their predictions of how they would respond in a different emotional situation. Specifically, when people are not emotionally aroused, they underestimate the impact of emotional arousal on their own preferences and behaviors. However, when people are emotionally aroused, they overestimate how much their preferences and behaviors in an unemotional situation state would resemble their current reactions.<sup>204</sup>

It is easy to see how empathy gaps could lead to inaccurate juror mindreading. The kinds of mental states at issue in criminal trials are quintessentially hot: heat of passion, insanity, duress, etc. Yet jurors in criminal trials are almost by definition in cold emotional states. Jurors should thus systematically underestimate the extent to which a defendant's act might have been influenced by emotion, imputing to the defendant a degree of rational calculation that he simply did not possess. Indeed, empathy gaps occur precisely because "people believe that their preferences and decisions are based primarily on the inherent desirability of choice alternatives rather than affectively influenced constructions of those alternatives."<sup>205</sup>

Consider, for example, an acquaintance rape case in which the defendant claims that the sex was consensual and the victim admits to having passionately kissed the defendant. That defense<sup>206</sup> requires jurors to determine how sexually aroused the victim was prior to the intercourse: did she want to have intercourse with the defendant, or did she want to stop with kissing? Because sexual arousal is a hot emotional state, cold jurors may well underestimate the potential effect of the consensual kissing on her overall level of sexual arousal—which means that they will be more likely to conclude that although she consented to the kissing, she would not have consented to the intercourse.

---

<sup>203</sup> Raymond S. Nickerson, *How We Know—and Sometimes Misjudge—What Others Know: Imputing One's Own Knowledge to Others*, 125 PSYCHOL. BULL. 737, 745 (1999).

<sup>204</sup> Van Boven & Loewenstein, *supra* note 73, at 49.

<sup>205</sup> Leaf Van Boven & George Loewenstein, *Empathy Gaps in Emotional Perspective Taking*, in OTHER MINDS: HOW HUMANS BRIDGE THE DIVIDE BETWEEN SELF AND OTHERS, *supra* note 54, at 284, 289 (citation omitted).

<sup>206</sup> Because nonconsent is an element of the crime in nearly every state, consent is technically a failure of proof defense, not an affirmative defense. See Paul H. Robinson, *Criminal Law Defenses: A Systematic Analysis*, 82 COLUM. L. REV. 199, 208 (1982).

The acquaintance rape case also requires jurors to mindread the defendant. How sexually desirous was he? Enough to force the victim to have intercourse if she wanted to stop with kissing? Here, too, empathy gaps may lead to inaccurate mindreading: cold jurors are likely to underestimate the level of force the defendant was willing to use to get what he wanted. Indeed, empirical evidence suggests that they do precisely that.<sup>207</sup>

The curse of knowledge is another mindreading error caused by the failure to quarantine. The curse refers to individuals' persistent tendency to let their knowledge interfere with their attempts to simulate others whom they know are less knowledgeable.<sup>208</sup> Numerous studies have demonstrated the existence of such curses,<sup>209</sup> including the financial forecasting experiment discussed earlier.

The existence of curses of knowledge has profound and unsettling implications for projection-based jury mindreading. Consider, for example, the classic mistake-of-fact defense in which a hunter claims that he cannot be convicted of murder because he honestly believed that his quarry was an animal, not a person. That defense depends on jurors' assessment of the hunter's knowledge when he pulled the trigger. Did he honestly believe he was shooting at an animal (in which case he must be acquitted, because the mistake negatives any subjective mens rea), or did he know full well the target was a person (in which case he should be convicted)? If the jurors fall prey to the curse of knowledge, they may well allow their *ex post* knowledge that the target was a person, not an animal, to contaminate their *ex ante* simulations of the defendant's perceptions immediately prior to the shooting, increasing the likelihood that they will conclude the defendant is lying.

Similar examples could be multiplied indefinitely, because *all* subjective mental states require jurors to assess the defendant's knowledge

---

<sup>207</sup> See George Loewenstein et al., *The Effect of Sexual Arousal on Expectations of Sexual Forcefulness*, 34 J. RES. IN CRIME & DELINQ. 443, 463 (1997) (“[I]ndividuals who were sexually aroused were more likely to imagine that they would behave in a sexually forceful manner on a date. The effect is evident not only in the comparison between the immediate-arousal and no-arousal conditions but also in the low anticipated levels of forcefulness in the prior-arousal condition.”). Loewenstein’s study focused on self-predictions, not social predictions. But as Van Boven and Loewenstein point out—and as indicated by projection research generally—“people in one emotional situation should project their current preferences and behaviors onto other people who are in different emotional situations.” Van Boven & Loewenstein, *supra* note 73, at 49.

<sup>208</sup> See, e.g., Nickerson, *supra* note 203, at 747 (“[T]he results from some studies suggest that people often overimpute their own knowledge to others; that is, they find it easy to impute to other people knowledge that they themselves have but others do not.”).

<sup>209</sup> See, e.g., GOLDMAN, *supra* note 50, at 166 (citing studies); Van Boven & Loewenstein, *supra* note 73, at 48 (citing studies).

at the time of the crime. A defendant only acts “recklessly,” for example, if “he consciously disregards a substantial and unjustifiable risk that the material element exists or will result from his conduct.”<sup>210</sup> Similarly, “willful blindness” exists “where it can almost be said that the defendant actually knew,” such as when the defendant “has his suspicion aroused but then deliberately omits to make further inquiries, because he wishes to remain in ignorance.”<sup>211</sup> Even “objective” mens rea requirements often involve knowledge determinations. As noted earlier, the MPC’s influential definition of negligence requires jurors to assess the risk overlooked by the defendant in light of “circumstances known to him,”<sup>212</sup> and many jurisdictions use subjectivized reasonableness standards that require the reasonableness of the defendant’s act to be assessed “seeing what (s)he sees and knowing what (s)he knows.”<sup>213</sup>

Indeed, criminal trials seem uniquely subject to the distorting effects of the curse of knowledge. Jurors who are asked to make knowledge determinations generally have at least a reasonably complete picture of the circumstances that actually existed at the time of the crime; the question they have to answer is how much less the defendant might have known—a determination that requires them to quarantine their knowledge of the “true” facts. Yet that is precisely the kind of quarantining that the curse of knowledge suggests is rarely cognitively possible.

#### b. Deficiency of Inputs

Even if jurors are able to quarantine *their own* idiosyncratic desires, beliefs, and knowledge from the projection process, they will still only be able to mindread accurately if they adjust their simulation routines to take into account *the defendant’s* idiosyncratic desires, beliefs, and knowledge. Such adjustment, which represents the theoretical component of projection,<sup>214</sup> requires jurors to complete two very different cognitive acts: (1) *recognizing* the need to adjust their simulations, and (2) *actually making* the necessary adjustments.

With regard to the first issue, it is clear that jurors do not *always* recognize the need to adjust their simulations.<sup>215</sup> Consider, for example, the mistake study discussed earlier. Finkel and his colleagues designed the

---

<sup>210</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(2)(c) (1985).

<sup>211</sup> WILLIAMS, *supra* note 44, at 159.

<sup>212</sup> MODEL PENAL CODE & COMMENTARIES § 2.02(d).

<sup>213</sup> *State v. Wanrow*, 559 P.2d 548, 557 (Wash. 1977) (citing *State v. Dunning*, 506 P.2d 321, 322 (Wash. 1973)).

<sup>214</sup> See *supra* notes 64-70 and accompanying text.

<sup>215</sup> See Goldman, *supra* note 178, at 83 (“People may not always take such factors into account.”).

study to determine whether mock jurors would be willing to “enter the defendant’s subjectivity completely” and acquit him regardless of the implausibility of his mistake—all four scenarios indicated that the subway vigilante honestly believed that he was about to be killed.<sup>216</sup> As it turned out, they weren’t: Finkel and his colleagues found that “[m]istake of fact works to lower culpability, but this mitigating effect depends on how believable or reasonable the mistake is and whether reasonable effort was made . . . despite the mistake.”<sup>217</sup> Those findings indicate that instead of recognizing the need to take the defendant’s psychological differences into account when they simulated his perceptions of the four self-defensive situations, jurors simply asked whether *they*—an unadjusted *they*—would have made the same mistakes. Hence the more implausible the mistake, the less willing they were to acquit.

Despite the mistake study, it seems clear that jurors who mindread through projection will *usually* recognize the need to take the defendant’s psychological differences into account. That is the conclusion Epley and his colleagues drew from a series of projection experiments they conducted:

Initially, people anchor on their own perspective, presumably because it is often highly accessible. In the absence of clear evidence that another person will see the world differently, perspective taking ends there with people assuming that others will perceive the world as they do. However, when others are known to be in different situations, from different backgrounds, or in possession of different knowledge, such *naïve realism* is simply untenable. In these cases, as in the experiments reported in this article, it is clear from the outset that one’s own perspective is not shared and adjustment is required.<sup>218</sup>

The critical question, then, is whether jurors can *actually* make the necessary adjustments to their egocentric simulations. Here Epley and his colleagues are more skeptical. Their research suggests that individuals project through an “anchor-and-adjust” process in which they “jump” further away from their own perspective until they reach a plausible reconstruction of the target’s perspective.<sup>219</sup> The problem is that, although the process reduces egocentrism,<sup>220</sup> mindreaders rarely jump far enough:

---

<sup>216</sup> Norman J. Finkel & Jennifer L. Groscup, *When Mistakes Happen: Commonsense Rules of Culpability*, 3 PSYCHOL. PUB. POL. & L. 65, 79 (1997). The defendant was not charged with negligent homicide, so even an honest but unreasonable mistake legally entitled him to an acquittal.

<sup>217</sup> *Id.* at 92.

<sup>218</sup> Epley et al., *supra* note 68, at 337 (citation omitted); *see also* Van Boven & Loewenstein, *supra* note 73, at 45 (“Mature adults typically realize that other people who are in a different situation may have beliefs, perceptions, preferences, and behavioral inclinations that differ from their own.”).

<sup>219</sup> Epley et al., *supra* note 68, at 337; *cf.* Hans Herbert Kogler, *Empathy, Dialogical Self, and Reflexive Interpretation: The Symbolic Source of Stimulation*, in EMPATHY AND

Although the anchoring and adjustment heuristic is clearly useful, the adjustments made from one's own perspective . . . tend to be insufficient—and give rise to egocentric biases—because they terminate once a plausible estimate is reached. This “satisficing” produces responses biased toward the egocentric side of the distribution of plausible estimates.<sup>221</sup>

This conclusion is also supported by jury decision-making research. Haney and Sweeney, for example, conducted a “simulation study of race and death sentencing” to determine why white jurors are far more likely to sentence black defendants to death than white defendants—an effect found in numerous death-penalty studies.<sup>222</sup> They found that white mock jurors “were significantly less willing or able to verbalize the reasons for their verdict choices” when the defendant and the victim were both black, referred more often and more positively to mitigating circumstances when the defendant was white, were far less likely to consider previous drug use to be a mitigating factor when the defendant was black, and were more likely to cite the absence of a criminal record as a reason for mercy when the defendant was white.<sup>223</sup> According to Haney, “[t]his suggests the racially discriminatory pattern of death sentencing may be a function of jurors’ inability to empathize with or enter the subjective world of . . . defendants who are racially different from them.”<sup>224</sup> That explanation is consistent with the anchor-and-adjust process, which would predict that white jurors would find it much easier to adjust their egocentric anchors to simulate the subjective perceptions of a sociologically similar defendant than a sociologically dissimilar one. As Harris says, “[t]he gist of the account is that simulation is more or less difficult depending on the number of adjustments that have to be made in default settings.”<sup>225</sup>

The problem is that jurors simply have no reliable method for determining how sociological differences translate into psychological differences, and thus have no reliable method for distinguishing between plausible and implausible adjustments.<sup>226</sup> How does a white juror

---

AGENCY: THE PROBLEM OF UNDERSTANDING IN THE HUMAN SCIENCES 194, 213 (Hans Herbert Kögler & Karsten R. Stueber eds., 2000) (“In order to begin the interpretive process, we have to project, of course, some understanding (naturally our own) onto the other’s expression and behavior. Yet such an interpretive projection can be nothing but a bridgehead into the other’s perspective, which needs to be continuously checked, revised, reassessed, and replaced as the interpretive process unfolds.”).

<sup>220</sup> Epley et al., *supra* note 68, at 337.

<sup>221</sup> *Id.* at 328 (citation omitted).

<sup>222</sup> See Haney, *supra* note 97, at 330.

<sup>223</sup> *Id.* at 331.

<sup>224</sup> *Id.*

<sup>225</sup> Harris, *supra* note 177, at 214.

<sup>226</sup> See Goldman, *supra* note 178, at 83 (noting that individuals “frequently . . . lack information to make adjustments”).

determine how the defendant's "blackness" affects his perceptions? How does a man determine how a battered woman would perceive a threatening situation? This is, of course, "the classic hermeneutic challenge[,]""<sup>227</sup> and there is no evidence that jurors can overcome it. On the contrary, "we have to understand that the holistic and culturally, socially, and historically impregnated nature of the background [of another] rules out any class or phenomena that could provide genuine and 'direct' access to the other's meaning."<sup>228</sup>

How, then, does a juror determine what adjustments he needs to make to his egocentric simulation of the defendant's situation? The most likely answer—through *stereotypes* of how different races, genders, etc. perceive and act in crime-related situations—is cause for concern.<sup>229</sup> Such stereotypes are all too common—blacks are violent, aggressive, and hostile,<sup>230</sup> women who dress provocatively want to have sex;<sup>231</sup> and so on. Stereotypes of that sort are almost always inaccurate; jurors that use them to adjust their simulations are likely to make even *less* accurate mental-state attributions than jurors who do not adjust their simulations at all. Natural similarity between jurors and defendants may be a myth, but there is no question that different sociological groups are more similar than stereotypes suggest.

This explanation presumes, of course, that jurors who adjust their simulations through stereotypes do not see the defendant as so dissimilar that projection is no longer possible. Adjustment has its limits: "it is reasonable to assume that people engage in such anchoring and adjustment only when it is clear that one's self is a reasonable starting place in estimating others' perspectives. In situations where one's own perspective is clearly irrelevant, [a] quite different process[] may be involved."<sup>232</sup> That process is *prototyping*, to which we now turn.

---

<sup>227</sup> Kogler, *supra* note 219, at 201.

<sup>228</sup> *Id.* at 213.

<sup>229</sup> See, e.g., Gordon, *supra* note 157, at 65-66 (suggesting that when an individual attempts to simulate the perceptions of someone from an "alien culture," he may begin with "a stereotypic set of adjustments").

<sup>230</sup> Jon Hurwitz & Mark Peffley, *Public Perceptions of Race and Crime: The Role of Race and Stereotypes*, 41 AM. J. POL. SCI. 375, 380 (1997).

<sup>231</sup> Malgorzata Wysocka-Pleczyk, *Victims of Violence: Stereotypes and the Process of Helping*, in UNDERSTANDING PROBLEMS OF SOCIAL PATHOLOGY 199, 203 (Przemyslaw Piotrowski ed. 2006).

<sup>232</sup> Epley et al., *supra* note 68, at 337.

## B. THE ERRORS OF PROTOTYPING

Unfortunately, prototyping is an even more problematic mindreading technique than projection. First, the features of prototypes rarely correspond to the legal definitions of particular crimes and defenses. Second, the fuzzy nature of prototyping is inconsistent with the formal requirements of jury decision-making. Third, prototypes tilt toward the extraordinary and unusual, reducing the likelihood that they will match actual cases. Fourth, prototypes vary significantly between jurors, undermining consistent decision-making. And fifth, prototyping in general causes jurors to process evidence in biased ways, leading to inaccurate fact-finding.

1. *Mismatched Elements*

One critical problem with prototyping is that the features of “people’s naive representations”<sup>233</sup> often deviate—sometimes substantially—from the legal elements of particular crimes and defenses. Some of those deviations focus on the criminal act itself. For example, 73% of the participants in Smith’s classic study of lay prototypes incorrectly believed that robbery requires the perpetrator to be armed.<sup>234</sup> Many deviations, however, focus on the mental states particular crimes require.

## a. Omitted Mens Rea

Some crime prototypes simply don’t include a mens rea requirement, thus implicitly holding the defendant strictly liable for committing the act. Stalking is an excellent example. As of 1998, forty-six states required stalking to be conducted either intentionally or purposefully.<sup>235</sup> A prototype study conducted by Dennison and Thomson, however, found that a stunning 95% of the 540 participants involved in the study classified harassing behavior<sup>236</sup> as stalking even when they believed “that the accused did not intend to cause fear or harm to the target.”<sup>237</sup>

---

<sup>233</sup> Smith, *supra* note 102, at 509; *see also* Cantor et al., *supra* note 103, at 46-47 (“[K]nowledge about any given category is structured around, and represented in long-term memory as, a *prototype* which captures the meaning of the category . . . . [A prototype may take the form of] an abstract set of features commonly associated with members of a category, with each feature assigned a weight according to degree of association with the category.”).

<sup>234</sup> Smith, *supra* note 104, at 861 tbl.1. Robbery actually requires only violence or intimidation. *See* BLACK’S LAW DICTIONARY 1354 (8th ed. 2004) (defining robbery). Smith’s mock jurors conflated robbery and armed robbery.

<sup>235</sup> *See* Susan M. Dennison & Donald M. Thomson, *Identifying Stalking: The Relevance of Intent in Commonsense Reasoning*, 26 LAW & HUM. BEHAV. 543, 544 (2002).

<sup>236</sup> Harassing behavior includes persistent following, making phone calls and hanging up,

Many of the crime prototypes in Smith's study are similarly deficient. Although assault requires the victim to have "a reasonable apprehension of imminent harmful or offensive contact,"<sup>238</sup> none of the assault prototypes held by mock jurors included that, or any other, mens rea.<sup>239</sup> Only 33% of their burglary prototypes included burglary's requirement<sup>240</sup> that the perpetrator intend to commit a felony.<sup>241</sup> Less than half of their murder prototypes contained any kind of intent requirement at all.<sup>242</sup>

#### b. Incorrect Mens Rea

Other crime prototypes contain a different flaw: they contain a mens rea requirement, but not the one legally required by the crime. Five percent of the participants in Dennison and Thomson's study, for example, recognized that stalking required mens rea. Unfortunately, all but one of that five percent (99.6%) believed that a defendant was guilty of stalking if he "should have realised that his actions were likely to create fear or apprehension in the target"<sup>243</sup>—a simple negligence standard.

Jurors also often make an even more problematic mistake: substituting *motive* for *intent*. The prosecution is never required to prove that the defendant had a motive for the crime;<sup>244</sup> a killer who intentionally kills for no reason is still guilty of murder. Nevertheless, Finkel and Groscup have found that juror prototypes for burglary, kidnapping, rape, and murder overwhelmingly emphasize motive<sup>245</sup>: motive not only "most dramatically differentiates the not guilty from the guilty,"<sup>246</sup> "when no . . . motive can be found, and particularly no guilty motive, then the prosecution's case collapses."<sup>247</sup> A similar transposition of mental elements occurs in insanity cases that involve a male defendant who kills a male stranger: "it is neither the criminal act, the precipitating event, nor even the defendant's psychiatric history that differentiates successful from unsuccessful outcomes, but motive."<sup>248</sup>

---

watching the victim's house, etc. See Dennison & Thomson, *supra* note 109, at 165.

<sup>237</sup> *Id.* at 166.

<sup>238</sup> BLACK'S LAW DICTIONARY, *supra* note 234, at 122 (defining assault).

<sup>239</sup> Smith, *supra* note 104, at 861 tbl.1.

<sup>240</sup> *Id.*

<sup>241</sup> BLACK'S LAW DICTIONARY, *supra* note 234, at 211 (defining burglary).

<sup>242</sup> Smith, *supra* note 104, at 861 tbl.1.

<sup>243</sup> Dennison & Thomson, *supra* note 109, at 167.

<sup>244</sup> See 21 AM. JUR. 2D *Criminal Law* § 128 (2008).

<sup>245</sup> Finkel & Groscup, *supra* note 110, at 224.

<sup>246</sup> *Id.*

<sup>247</sup> *Id.* at 225.

<sup>248</sup> *Id.* at 219.

## c. Redefined Mens Rea

Finally, some prototypes include the “correct” mens rea, but understand that mental state in a legally inappropriate manner. The insanity defense is perhaps the best example: numerous studies have shown that “jurors construe case facts and render verdicts” on the basis of “complex, multifaceted prototypes of insanity that cannot be reduced to legal tests of insanity or to psychiatric diagnoses.”<sup>249</sup> Skeem and Golding have found, for example, that jurors generally possess one of three basic insanity prototypes: a “severe mental disability” prototype, which involves “a peculiar blend of psychosis and intellectual impairment” that renders the insane defendant little more than a “wild beast” (47%);<sup>250</sup> a “moral insanity” prototype, which “conflates characteristics of psychopathy and psychosis to represent a malevolent, detached, irrational, and unpredictably violent offender” (33%);<sup>251</sup> and a “mental-state centered” prototype, which focuses almost exclusively on impairments in the defendant’s mental state at the time of the crime (21%).<sup>252</sup> All three prototypes strongly influence jurors’ willingness to find a defendant insane,<sup>253</sup> but only the third—held by little more than one in five jurors—is consistent with the current legal definition of insanity.<sup>254</sup>

Self-defense’s objective reasonableness requirement is another example of an often-redefined mens rea. As mentioned earlier, Russell and Melillo’s prototype study found that jurors were much more likely to acquit a “typical” battered woman than an “atypical” battered woman, even when the situation was objectively the same.<sup>255</sup> The reason why is revealing—jurors simply perceived the defendant’s act *as more objectively reasonable* when they considered her to be a typical battered woman:

Typical defendants were perceived to have more options available . . . . In contrast, respondents rated typical defendants as more likely to meet the requirements of self-defense. Respondents in passive conditions also perceived the defendant as having fewer options to killing and were most likely to find a fit with the typology of a

---

<sup>249</sup> Skeem & Golding, *supra* note 117, at 604.

<sup>250</sup> *Id.* at 593, 605; *see also* Roberts et al., *supra* note 138, at 222-23 (reporting that in their study of insanity prototypes, “bizarre crimes were perceived as indicating more mental disorder”).

<sup>251</sup> Skeem & Golding, *supra* note 117, at 601.

<sup>252</sup> *Id.* at 595, 601.

<sup>253</sup> *Id.* at 604.

<sup>254</sup> *Id.* at 604-05; *see also* Roberts et al., *supra* note 138, at 222-23 (“Case law cautions that bizarre and heinous actions should not be construed as an indication of either mental illness or insanity.”).

<sup>255</sup> Russell & Melillo, *supra* note 135, at 234 (“[P]articipants in atypical/active conditions were more likely to assign judgments of second-degree murder, whereas those in typical/passive conditions most often rendered not-guilty verdicts.”).

battered woman . . . . Typicality and response history also resulted in case effects. Respondents in passive conditions found the defendant was more likely to meet the requirements of self-defense [and] was more likely to fit the typology of a battered woman . . . .<sup>256</sup>

Jurors' personological prototypes for battered women, in other words, naturally<sup>257</sup> include a "subjectivized" reasonableness standard that views the reasonableness of the woman's act from her subjective perspective—"seeing what (s)he sees and knowing what (s)he knows." That prototype would be legally accurate in the bare majority of states that use a subjectivized self-defense standard,<sup>258</sup> but would be inaccurate everywhere else.

## 2. Fuzzy Prototyping

The disjunction between the features of prototypes and the legal elements of crimes and defenses is only part of the problem. Jurors also do not view prototype features as "singly necessary and jointly sufficient" for conviction: if the characteristics of the defendant's crime match the features of the prototype closely enough, jurors will convict even in the absence of one (or more) of the features,<sup>259</sup> a basic violation of due process.<sup>260</sup>

This typicality effect can easily lead to inaccurate mental-state attributions. Most important, it leaves open the possibility that a juror will convict a defendant whose criminal act closely matches his prototype for that kind of crime even if the defendant did not possess the necessary mens rea. Finkel and Groscup have found, for example, that although jurors will "usually" acquit a burglary defendant who breaks into a house he previously occupied to retrieve forgotten belongings, some jurors will still convict.<sup>261</sup> Such convictions are legally indefensible, because a defendant cannot intend to steal his own possessions. But they make sense as an artifact of prototyping. Although the "misunderstood burglar"<sup>262</sup> scenario does not match the mental element of the most common burglary prototype ("purpose is to steal"), it matches all three of that prototype's physical elements (a break-in, a house or apartment, the taking of something valuable).<sup>263</sup> The misunderstood burglar scenario is thus typical enough—at least for the jurors who convict—to warrant being considered burglary.

---

<sup>256</sup> *Id.*

<sup>257</sup> The jury instructions used in the experiment contained a purely objective standard.

<sup>258</sup> For a state-by-state background current to 1998, see Heller, *supra* note 173, app.

<sup>259</sup> See *supra* notes 118-23 and accompanying text.

<sup>260</sup> See, e.g., *Patterson v. New York*, 432 U.S. 197, 210 (1977).

<sup>261</sup> See Finkel & Groscup, *supra* note 110, at 223.

<sup>262</sup> The term is Finkel & Groscup's. See *id.*

<sup>263</sup> See Smith, *supra* note 104, at 861 tbl.1.

Typicality led to a similar error in Smith's study of murder prototypes. Smith presented mock jurors with a murder scenario in which the defendant intentionally induced a fatal heart attack in the victim by holding him in front of an oncoming train, knowing that the victim had a heart condition.<sup>264</sup> The scenario legally qualified as murder, and most jurors had little problem categorizing it as such.<sup>265</sup> Smith then presented different mock jurors with a version of the scenario in which it was impossible to find that the defendant acted with the necessary intent, because the defendant did not know about the victim's heart. Despite the legal insufficiency of the new scenario, 26% of the jurors were still willing to convict.<sup>266</sup> Smith thus concluded that although intentionality is a "relatively important" feature of most jurors' murder prototypes, it is not a necessary one—jurors will overlook it if a killing is otherwise sufficiently prototypical.<sup>267</sup>

Even if jurors do not completely ignore mens rea, the typicality effect can lead them to discount its importance. An experiment conducted by Follingstad and her colleagues asked mock jurors to judge three scenarios in which a battered woman killed her husband and claimed self-defense.<sup>268</sup> The scenarios were designed to test the jurors' willingness to view the woman's actions as objectively reasonable: in Scenario 1, the husband advanced on the woman with a weapon, and she used a weapon to kill him; in Scenario 2, the woman killed her husband after he advanced toward her, unarmed but making verbal threats; in Scenario 3, the husband beat and verbally threatened the woman, but she waited until he was asleep to kill him.<sup>269</sup> The results were surprising: although the jurors were far more likely to acquit the defendant in Scenario 1 than in Scenarios 2 and 3, nearly half of the jurors *still* acquitted her in the latter two scenarios—and there was no significant difference between them.<sup>270</sup>

The results are legally inexplicable; if the mock jurors had interpreted self-defense's imminence requirement "in a strictly objective fashion," Scenario 3 would have led to far fewer acquittals than Scenario 2.<sup>271</sup> The results make perfect sense, however, if the jurors determined whether to convict by matching the characteristics of the defendant's act to their

---

<sup>264</sup> *Id.* app. B (presenting murder).

<sup>265</sup> *Id.* at 867. Note, though, that 33% of the jurors still refused to convict, indicating that their murder prototypes did not track the legal definition of murder. *Id.*

<sup>266</sup> *Id.*

<sup>267</sup> *Id.*

<sup>268</sup> Diane R. Follingstad et al., *Factors Predicting Verdicts in Cases Where Battered Women Kill Their Husbands*, 13 LAW & HUM. BEHAV. 253, 259 (1989).

<sup>269</sup> *Id.*

<sup>270</sup> *Id.* at 266.

<sup>271</sup> See FINKEL, *supra* note 28, at 243.

prototypes of what the “typical” battered-woman self-defense situation looks like. The 50% of mock jurors who acquitted the defendant in Scenario 3, that is, might simply have viewed the circumstances of the crime—the husband’s violent history, the beating itself, the verbal threatening, etc.—as so typical of a legitimate defensive act by a battered woman that the objective unreasonableness of the woman’s belief regarding imminence did not affect their categorization of the killing as self-defense.

The typicality effect can also go the other direction: if a juror finds that the defendant acted with the mental state associated with the prototype of a particular crime, he may convict even if the defendant’s act does not otherwise satisfy the crime’s legal requirements. That is exactly what happened when Smith presented jurors with a robbery scenario in which the defendant took a cassette recorder he found in a park next to a tree.<sup>272</sup> The scenario did not legally qualify as robbery, because the defendant did not take the property from the owner by force or threat of force. Nevertheless, the scenario produced an 87% conviction rate, indicating that jurors so strongly associate<sup>273</sup> the intentional taking of property with robbery that the presence of intentional taking will trump the absence of the crime’s physical elements.<sup>274</sup>

### 3. *Extraordinary Prototypes*

We have seen that, when a juror mindreads through prototyping, mental-state attributions are a function of the degree to which the defendant’s act matches a juror’s prototype for a particular crime or defense: if the match is strong enough, the juror will attribute the mental state associated with the prototype to the defendant.<sup>275</sup> That explanation raises a critical question, one that I have largely glossed over to this point: if matching involves a comparison between the content of a prototype and the nature of the defendant’s act, what do we mean when talk about a prototype’s “content”? Are prototypes “pictures in the head,” to appropriate Walter Lippman’s classic description of stereotypes?<sup>276</sup> Are

---

<sup>272</sup> Smith, *supra* note 104, app. B (presenting robbery).

<sup>273</sup> See Cantor et al., *supra* note 103, at 47 (noting that each prototypical feature is “assigned a weight according to degree of association with the category”).

<sup>274</sup> Smith, *supra* note 104, at 867. This result cannot be explained by arguing that the mock jurors who convicted did not know about the other elements; 73% of the participants in an earlier experiment in Smith’s study listed force or the threat of force as a feature of robbery. See *id.* at 861 tbl.1.

<sup>275</sup> See, e.g., Skeem & Golding, *supra* note 117, at 569 (“The more closely the defendant’s attributes match those of her prototype, the more likely she is to judge that defendant a member of the category ‘insane.’”).

<sup>276</sup> WALTER LIPPMANN, PUBLIC OPINION 25 (1922).

they abstract sets of features, with no representational aspects? Or are they a combination of the two?

The answer is that it depends on the prototype. Some prototypes are, in fact, nothing more than “pictures in the head”—visual exemplars of what the typical crime or defense “looks like.” The prototypic assault, for example, is simply a physical attack in which the victim is injured. Matching that kind of prototype is thus a wholly imaginative process in which the juror accesses the visual prototype and compares it to the image of the defendant’s act that he has generated from the evidence in the case.<sup>277</sup> No theoretical comparison is required between the prototype and the act, because the prototype does not contain any abstract features.

But what about mental states? They are inherently abstract; you can’t see a mental state. So doesn’t a juror have to engage in theoretical reasoning to determine whether the defendant acted with the necessary mens rea?

There are two responses to that question. The first is that it is often moot, because some juror prototypes simply do not contain a mental element.<sup>278</sup> The more complicated response is that, when a purely visual prototype “contains” a mental element, the necessary mental state is *represented as a particular kind of action*. Many juror prototypes of consent in rape cases, for example, represent the mental state of “consent” as the absence of physical or verbal resistance to sex.<sup>279</sup> Similarly, many juror prototypes of insanity equate the mental state of being insane with looking and acting bizarrely.<sup>280</sup> With these kinds of prototypes, action serves as a proxy for the underlying mental state that (ostensibly) produced it; certain actions signify certain mental states.

Other prototypes are more complex, involving a combination of visual images and abstract features. A common prototype of a non-criminal burglary, for example, involves “a former roommate or ex-spouse return[ing] to a house he or she lived in to retrieve some forgotten belongings.”<sup>281</sup> Entering the house and taking the belongings are visual; the defendant owning the belongings is abstract. Similarly, the prototypical battered woman is young, fragile, guilt-ridden, confused, and depressed; rarely interacts with others; has children and is financially dependent on her

---

<sup>277</sup> For an explanation of how jurors generate images of the defendant’s act, see generally Kevin Jon Heller, *The Cognitive Psychology of Circumstantial Evidence*, 105 MICH. L. REV. 241-305 (2006).

<sup>278</sup> See *supra* notes 235-42 and accompanying text.

<sup>279</sup> See Littleton & Axsom, *supra* note 110, at 474.

<sup>280</sup> See Roberts et al., *supra* note 138, at 222-23.

<sup>281</sup> Finkel & Groscup, *supra* note 110, at 223.

partner; makes excuses for her partner's behavior; and passively accepts her abuse. Very few of those features are representational.

When a prototype contains both visual elements and abstract features, matching will be a *hybrid* process, both imaginative and theoretical. To apply the "misunderstood burglar"<sup>282</sup> prototype, for example, a juror will have to both (1) imaginatively compare the "entering the house" and "taking the belongings" features to the defendant's actions, and (2) theoretically compare the "ownership" feature to the evidence presented in the case—were the belongings really his? Similarly, to determine whether the defendant is a battered woman, a juror will visually compare the representational features of his "battered" prototype to the defendant's appearance (young, trying to hide injuries, depressed, etc.) and theoretically compare the non-representational features to the evidence in the case (rarely interacts with others, financially dependent, makes excuses, etc.).

Mindreading through complex prototypes can involve either visual proxies or theoretical comparison. Consider what Finkel calls the "benevolent burglar" prototype: a poor defendant who steals food and medicine from a store because his children are sick and need to eat.<sup>283</sup> That prototype is a visual/abstract hybrid: the defendant stealing the items is visual; the defendant's motive (to take care of his hungry and sick kids) is abstract. Some jurors may match the defendant's abstract motive through a visual proxy, interpreting stealing food and medicine (but nothing else) as signifying a benevolent motive. Other jurors, by contrast, may match that motive theoretically, by examining whether the evidence in the case indicates that the defendant's children needed food and medicine.<sup>284</sup>

Regardless of whether a prototype is simple or complex, it is the content of the prototype that drives the matching process. Prototyping can thus lead to inaccurate mindreading *even when* the structural features of a juror's prototype do, in fact, match the legal elements of a crime or defense: namely, when the prototype's visual or abstract content is so idiosyncratic that legitimate instances of the particular crime or defense will fail to match it. Put more simply, prototyping will be inaccurate whenever the juror's idea of the "typical" crime or defense isn't actually all that typical.

Unfortunately, research indicates that prototypes are rarely representative of the crime or defense they purport to represent. Finkel and Groscup, for example, found that "against the backdrop of typical legal cases and empirical studies of actual euthanasia, insanity, heat of passion,

---

<sup>282</sup> *Id.*

<sup>283</sup> *Id.*

<sup>284</sup> A benevolent motive, of course, does not mean that the defendant lacked the mens rea of burglary. Finkel and Groscup's research indicates, however, that jurors are nevertheless likely to acquit the benevolent burglar. *Id.*

and self-defense cases, the *extraordinary* emerges pervasively and dramatically in the prototypes.”<sup>285</sup> More specifically, Stalans and Diamond have concluded that many jurors “have a picture of the typical burglary case that is more severe than the burglary cases that normally appear in court.”<sup>286</sup> Olsen-Fulero and Fulero found in their study of rape prototypes that “[v]irtually all participants described a rape by a stranger occurring at night in a park or parking lot, even though most real rapes occur in a woman’s own home or during a date,” suggesting that “a juror’s commonsense schema of rape is very often discrepant from the facts of a given or typical court case.”<sup>287</sup> And Skeem and Golding have found that the typical defendant who is acquitted by reason of insanity—a psychotic man with a psychiatric history who commits a violent crime—matches only 21% of juror insanity prototypes, suggesting that “many jurors’ prototypes of insanity are prone to convict.”<sup>288</sup>

The culprit, not surprisingly, is the media: television, the movies, and the news. “Because a relatively small proportion of the public has direct experience with the justice system, public knowledge and views of law and the legal system are largely dependent on media representations.”<sup>289</sup> I have already mentioned Holst and Pezdek’s study of robbery prototypes, which found that nearly 90% of the participants had learned what a typical robbery looked like from the media.<sup>290</sup> Those results are by no means unusual. When Finkel and Groscup conducted content analyses of the “extraordinary” prototypes in their study, they found that the prototypes “reveal[ed] obvious news media, TV, and Hollywood influences, where the accessibility bias is quite evident.”<sup>291</sup> Similarly, Skeem and Golding attributed the two inaccurate insanity prototypes held by 79% of their subjects—the “wild beast” prototype (47%) and the “moral insanity” prototype (32%)—to media representations of insanity.<sup>292</sup>

---

<sup>285</sup> *Id.* at 216.

<sup>286</sup> Loretta J. Stalans & Shari Seidman Diamond, *Formation and Change in Lay Evaluations of Criminal Sentencing: Misperception and Discontent*, 14 LAW & HUM. BEHAV. 199, 205 (1990).

<sup>287</sup> Olsen-Fulero & Fulero, *supra* note 81, at 404.

<sup>288</sup> Skeem & Golding, *supra* note 117, at 605; *see also* Valerie P. Hans & Dan Slater, “Plain Crazy:” *Lay Definitions of Legal Insanity*, 7 INT’L J.L. & PSYCHIATRY 105, 111 (1984) (“[I]n defining insanity, some of our respondents insisted on near total lack of comprehension. Very few, if any, defendants would be able to meet such a strict standard.”).

<sup>289</sup> Hans, *supra* note 140, at 399.

<sup>290</sup> Holst & Pezdek, *supra* note 141, at 580.

<sup>291</sup> Finkel & Groscup, *supra* note 110, at 216.

<sup>292</sup> *See* Skeem & Golding, *supra* note 117, at 592 (“The [Severe Mental Disability]’s characteristics are identifiable not only in research, but also in contemporary media and historical representations of insanity. The news and entertainment media are laypeople’s

#### 4. *Prototype Diversity*

Prototype diversity exists at two levels. First, an individual juror may possess multiple prototypes of the same crime or defense. Holst and Pezdek, for example, have shown that jurors normally have at least three different robbery prototypes: one for a bank robbery, one for a convenience store robbery, and one for a mugging.<sup>293</sup> Similarly, Wiener and his colleague's study of murder prototypes found that "[a] careful reading of the story skeletons . . . evidenced multiple (at least three and maybe five) shared prototypes."<sup>294</sup> Indeed, it even possible—perhaps likely—that jurors possess “guilty” and “not guilty” versions of each prototype.<sup>295</sup> Participants in Finkel and Groscup's prototype study had little difficulty generating prototypes of burglaries, kidnappings, rapes, and killings in self-defense that required acquittal.<sup>296</sup>

This kind of diversity is unproblematic. In fact, it is almost certainly a good thing, because the existence of multiple prototypes will increase the likelihood that real-world crimes and defenses will match a particular prototype. The problem with a single prototype (and particularly an extraordinary one) is that the gap between the prototype and the set of legitimate instances of the crime or defense is simply too large: too few instances will overlap with the prototype enough to qualify as a match. That gap is obviously much smaller if a juror possesses multiple prototypes of a particular crime or defense, because more legitimate instances will match (one of) the prototypes. And that is true even if the multiple prototypes are themselves skewed toward the extraordinary, which seems likely.

Prototypes also normally differ *between* jurors. Skeem and Golding's study, for example, revealed not only that jurors possess three different insanity prototypes, but also that “jurors do not substantially agree on even a subset of the features that characterize insanity.”<sup>297</sup> Such differences are not accidental; a similar study by Hans and Slater found that “the demographic variables of education, gender, and race, and television viewing were all related to the way people defined legal insanity.”<sup>298</sup>

---

primary sources of information about psychopathology and the insanity defense.”); *id.* at 594 (“Most striking . . . is the resemblance of the [Moral Insanity] prototype to the most persistent, pervasive, and powerful media stereotype of the mentally ill: the psychotic psychopath.” (citations omitted)).

<sup>293</sup> Holst & Pezdek, *supra* note 141, at 578.

<sup>294</sup> Wiener et al., *supra* note 124, at 135.

<sup>295</sup> See Finkel & Groscup, *supra* note 110, at 213.

<sup>296</sup> *Id.* at 225.

<sup>297</sup> Skeem & Golding, *supra* note 117, at 582.

<sup>298</sup> Hans & Slater, *supra* note 288, at 111.

Crime prototypes also exhibit between-juror differences. Smith's study of assault, burglary, kidnapping, murder, and robbery,<sup>299</sup> for example, found that "[a]cross the five crimes, very few features were listed consistently by subjects. No feature was listed by all subjects, and only one or two features for each crime were provided by as many as half of the subjects."<sup>300</sup> Wiener and his colleagues reached a similar conclusion in their murder study: "[N]o single- or even multiple-murder story prototypes emerged. In fact, the amount of variability among the stories on their faces is striking."<sup>301</sup>

Finkel and his colleagues' infanticide study provides a particularly dramatic example of this variability. As noted earlier, mock jurors generally agreed that beating the baby to death signified premeditation, while suffocation signified a temporary loss of reason. Their prototypes for abandonment, however, were very different. Some jurors equated abandoning the baby in the dumpster with the absence of premeditation, saying things like "[she] didn't kill," and "the baby could have been found." Others, however, drew exactly the opposite conclusion; in their view, because "she took a long walk" and "had time to think and stop," she "had a plan," was "trying to kill," and acted with "malice" and "premeditation."<sup>302</sup>

Unlike within-juror diversity, between-juror diversity *is* problematic. Juries normally employ one of two different deliberation styles. The first is "evidence-driven," in which jurors begin by trying to reach a consensus about what actually happened during the crime and then attempt to agree on the appropriate verdict for the consensus story. The second is a "verdict-driven," in which jurors begin by taking a public vote on the defendant's guilt and then attempt to work out their differences by discussing the evidence in the case.<sup>303</sup>

Both deliberation styles will be more difficult when jurors bring different prototypes of the relevant crime or defense into the jury room. A jury that uses an evidence-driven style will find it particularly difficult to agree on the appropriate verdict, because the jurors will not be using the same legal standards to assess the consensus story. Even worse, those jurors will most likely not even *realize* they are applying different legal

---

<sup>299</sup> *But cf.* Holst & Pezdek, *supra* note 141, at 578 (finding, in their study of robbery prototypes, "a high rate of agreement among subjects for the actions that comprise each type of robbery").

<sup>300</sup> Smith, *supra* note 104, at 860.

<sup>301</sup> Wiener et al., *supra* note 124, at 129.

<sup>302</sup> Finkel et al., *supra* note 113, at 1129-30.

<sup>303</sup> See Nancy Pennington & Reid Hastie, *Practical Implications of Psychological Research on Juror and Jury Decision Making*, 16 PERSONALITY & SOC. PSYCHOL. BULL. 90, 99-100 (1990).

standards: although the judge will have instructed them on the “correct” definitions of the relevant crimes or defenses, creating the appearance of uniformity, their verdict choices will still be determined by their prototypes.<sup>304</sup> As a result, the jury will almost always reach the verdict that is consistent with whatever prototype happens to be held by the most jurors, because numerous studies have shown that the best predictor of a jury’s final verdict is the distribution of its initial verdict preferences.<sup>305</sup>

A jury that uses a verdict-driven style will have the same problem, but to an even greater degree. “[W]hen groups begin discussion with a clear preference, discussion tends to strengthen those opinions—a phenomenon called ‘group polarization.’”<sup>306</sup> By beginning with a vote, in other words, a verdict-driven jury will “lock in” the initial distribution of (prototype-driven) verdicts, making it very unlikely that subsequent discussion of the evidence in the case will affect verdict preferences. At least with the evidence-driven jury, the attempt to create a consensus story may affect how some jurors view the evidence. If it does, the jury’s initial distribution of verdict preferences will likely be different than the verdict-driven jury’s, because a change in how a juror views the evidence in a case can affect a juror’s initial verdict preference. This can happen either by potentially reducing the match between the evidence and the juror’s prototype to the point that the prototype no longer applies (or vice-versa), or by triggering a different prototype of the same crime, one that leads to a different verdict preference.

### 5. Biased Evidence Processing

The final problem with prototyping is that it will often lead jurors to inaccurately assess the evidence in a defendant’s case. That is a critical flaw, because inaccurate evidence assessment will then lead to inaccurate matching—either increasing the likelihood of a false match with the juror’s prototype or decreasing the likelihood of a true match.<sup>307</sup>

---

<sup>304</sup> See, e.g., Smith, *supra* note 104, at 866 (“These results indicate that hearing the judge’s instructions on the crime definitions did not alter the strategies subjects used to make their verdict decisions. Typicality still influenced categorization for assault, burglary, and kidnapping, even when subjects were instructed in the legally correct decision strategy.”).

<sup>305</sup> See, e.g., H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* 488 (1966) (finding in their study that the majority pre-deliberation verdict became the final verdict in 209 out of 215 juries, or 97%).

<sup>306</sup> NEIL BREWER & KIPLING D. WILLIAMS, *PSYCHOLOGY AND LAW: AN EMPIRICAL PERSPECTIVE* 368 (2005) (citation omitted).

<sup>307</sup> See Vicki L. Smith & Christina A. Studebaker, *What Do You Expect?: The Influence of People’s Prior Knowledge of Crime Categories on Fact-Finding*, 20 *LAW & HUM. BEHAV.* 517, 529 (1996).

Two errors are particularly important. A juror who initially believes that the defendant's act matches his prototype of a particular crime or defense will likely (1) exhibit a significant confirmation bias in his evidence assessment, and (2) misremember the evidence presented at trial in ways that increase the extent of the match, what cognitive psychologists call "gap filling."

a. The Confirmation Bias

To accurately match evidence to a prototype, a juror must be able to distinguish objectively between evidence that supports a match and evidence that does not. Unfortunately, once jurors have reached an initial conclusion about whether a match exists, they will find it very difficult to recognize evidence that is inconsistent with that conclusion.<sup>308</sup> First, they may limit their search for such evidence, focusing almost exclusively on evidence that supports the match.<sup>309</sup> That is exactly what Jones and Kaplan found in their study of race-crime prototypes:

A confirmatory strategy is used in examining information about White and Black defendants charged with a racially congruent crime because the convergence of race with the crime stereotype suggests a plausible hypothesis of guilt. This hypothesis, in turn, seems to lead to a greater desire for evidence that confirms it.<sup>310</sup>

Second, jurors may ignore questions about the reliability of evidence that supports the match, a form of the confirmation bias known as "belief perseverance": the tendency of individuals to "adhere to their beliefs when the original evidential basis of the beliefs is shown to be flimsy, false, or nonexistent."<sup>311</sup> We have seen, for example, that the existence of a guilty motive is the most heavily weighted feature of many crime prototypes. A juror who believes that the defendant had a motive to commit the crime will thus be very likely to overlook problems with the evidence on which that belief is based—it is hearsay, the witness who established the motive had reason to lie, etc. Indeed, research indicates that belief perseverance is strongest when the discredited evidence is part of a "coherent, causally related account in which a single or minimal correction has a significant impact on the construal of meaning."<sup>312</sup> That is exactly the situation with a

---

<sup>308</sup> See, e.g., David M. Sanbonmatsu et al., *Overestimating Causality: Attributional Effects of Confirmatory Processing*, 65 J. PERSONALITY & SOC. PSYCHOL. 892, 893 (1993).

<sup>309</sup> See, e.g., David M. Sanbonmatsu et al., *The Subjective Beliefs Underlying Probability Overestimation*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 276, 278 (1997).

<sup>310</sup> Jones & Kaplan, *supra* note 128, at 10.

<sup>311</sup> Martin F. Davies, *Belief Persistence After Evidential Discrediting: The Impact of Generated Versus Provided Explanations on the Likelihood of Discredited Outcomes*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 561, 562 (1997).

<sup>312</sup> Hollyn M. Johnson & Colleen M. Seifert, *Sources of the Continued Influence Effect*:

prototype match involving a heavily weighted feature like motive—rejecting the match will require rejecting the “coherent, causally related” prototype itself.

Third, and finally, jurors may interpret ambiguous evidence in a manner consistent with the prototype match.<sup>313</sup> For example, if the evidence strongly, but not definitively, suggests that the defendant did not have a motive to commit the crime, jurors may give the prosecution the benefit of the doubt and still assume that the defendant had one.

#### b. Gap-Filling

Even more troubling, a number of studies have demonstrated that a prototype match may lead jurors to misremember the evidence presented at trial, filling gaps in the evidence in ways that are consistent with the prototype. Holst and Pezdek’s study of robbery prototypes is an excellent example. According to their research, the most common prototype of a mugging contains the following sequence of acts by the defendant:

- Make a plan.
- Look around.
- Select a victim.
- Wait for an opportunity.
- Approach the victim.
- Use force.
- Take money/valuables.
- Leave the scene.<sup>314</sup>

To determine whether gap-filling takes place, Holst and Pezdek gave college students the transcript of a mock mugging in which a witness testified that the defendant had taken all of those prototypical actions except one—using force. They then asked the students a week later to recall as much of the witness testimony as they could. More than 20% of the students erroneously recalled the witness testifying that the defendant used force.<sup>315</sup> Moreover, the unstated action was “apparently so real in their memory that they were actually significantly more confident deciding that [the action] had been stated than not.”<sup>316</sup>

---

*When Misinformation in Memory Affects Later Inferences*, 20 J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY & COGNITION 1420, 1432 (1994).

<sup>313</sup> See Derek J. Koehler, *Explanation, Imagination, and Confidence in Judgment*, 110 PSYCHOL. BULL. 499, 511 (1991).

<sup>314</sup> Holst & Pezdek, *supra* note 141, at 580 tbl.3.

<sup>315</sup> *Id.* at 583.

<sup>316</sup> *Id.* at 585.

Lenton and Bryan reached a similar conclusion in their study of sexual intent attributions. They presented college students with dating situations that previous research indicated were prototypic of either an interest in casual sex or a committed relationship, but omitted certain central features of those situations. Fifteen minutes later, the students completed a recognition test in which they attempted to recall what had happened between the man and the woman. Ninety-eight percent of the participants made at least one script-consistent error, and 41% made at least *four*.<sup>317</sup>

### VIII. DEBIASING PROJECTION AND PROTOTYPING

For all the reasons discussed in the previous Part, neither projection nor prototyping will consistently lead to accurate mindreading. This is a depressing conclusion, and one that raises an obvious question: what should we do about those techniques?

It is clear that we cannot *prevent* jurors from projection and prototyping. Jurors project “even when they are asked not to,”<sup>318</sup> “regardless of information they have about other individuals,”<sup>319</sup> and “even when they have unambiguous information that their own behavior is uninformative.”<sup>320</sup> Similarly, jurors not only prototype even when they are warned of the dangers of prototyping,<sup>321</sup> jury instructions defining crimes and defenses have absolutely no impact on the content of their prototypes.<sup>322</sup> In fact, asking jurors not to prototype will almost certainly *backfire*, making the prototypes even more salient<sup>323</sup> and—worse—increasing the likelihood that the jurors will falsely remember prototype-consistent evidence.<sup>324</sup>

---

<sup>317</sup> Alison P. Lenton & Angela Bryan, *An Affair to Remember: The Role of Sexual Scripts in Perceptions of Sexual Intent*, 12 PERS. RELATIONSHIPS 483, 492 (2005).

<sup>318</sup> Clement & Krueger, *supra* note 72, at 219.

<sup>319</sup> *Id.*

<sup>320</sup> Van Boven & Loewenstein, *supra* note 73, at 59.

<sup>321</sup> Smith, *supra* note 102, at 532 (“[T]his supplementary instruction had no effect on decision making; subjects relied just as heavily on their prior knowledge of the target crime.”).

<sup>322</sup> *See, e.g.*, Smith, *supra* note 104, at 866 (“These results indicate that hearing the judge’s instructions on the crime definitions did not alter the strategies subjects used to make their verdict decisions. Typicality still influenced categorization for assault, burglary, and kidnapping, even when subjects were instructed in the legally correct decision strategy.”).

<sup>323</sup> *See, e.g.*, Richard M. Wenzlaff & Daniel M. Wegner, *Thought Suppression*, 2000 ANN. REV. PSYCHOL. 59, 61.

<sup>324</sup> *See, e.g.*, Maarten J.V. Peters et al., *When Stereotypes Backfire: Trying to Suppress Stereotypes Produces False Recollections of a Crime*, 11 LEGAL & CRIM. PSYCHOL. 327, 328 (2006) (“False stereotype-consistent information is especially likely to be reported when people are instructed to deliberately suppress stereotypes.”).

The best we can hope for, then, is to improve the accuracy of projection and prototyping. This Part discusses the most promising debiasing techniques.

#### A. DEBIASING PROJECTION

Three techniques exist to minimize projection's inherent egocentrism. First, empathy gaps can be reduced by asking individuals in a cold emotional state to imagine how they would think and feel in a hot emotional state before they predict how they would react in a hot state<sup>325</sup>—the imagination task apparently makes the individuals realize that they are likely to underestimate the differences between their hot and cold states, leading them to take those differences into account when they self-predict. That is an encouraging finding, because it suggests that asking jurors to imagine themselves in hot emotional states prior to mindreading may encourage them to take the defendant's hot emotional state into account when they simulate his experience of the crime. As Van Boven and Loewenstein point out, "biased predictions of their own feelings play an important role in producing biased predictions of others' feelings."<sup>326</sup>

A similar technique may help jurors avoid falling prey to the curse of knowledge—a mindreader's failure to consider the possibility that the target knows less than he does. Research indicates that the curse can be minimized by having the mindreader recall a situation in which he was less knowledgeable than he is now before he determines how much the target knows.<sup>327</sup> The imagination task seems to sensitize the mindreader to the possibility that he may know more than the target, leading him to more accurately reconstruct the state of the target's knowledge. It may thus be useful to have jurors engage in the same kind of imagination task prior to beginning their deliberations, or even prior to hearing evidence.

Finally, individuals project less egocentrically if they are provided financial incentives to accurately mindread.<sup>328</sup> Adjusting projection requires effort<sup>329</sup> and individuals try to conserve their cognitive resources whenever possible,<sup>330</sup> so individuals who have an incentive to adjust accurately are more likely to exert the necessary effort than individuals who

---

<sup>325</sup> See Loewenstein et al., *supra* note 207, at 466.

<sup>326</sup> Van Boven & Loewenstein, *supra* note 73, at 52.

<sup>327</sup> See, e.g., Pamela J. Hinds, *The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance*, 5 J. EXPERIMENTAL PSYCHOL., Appendix 1999, at 205, 207.

<sup>328</sup> See, e.g., Epley et al., *supra* note 68, at 337.

<sup>329</sup> *Id.* at 332.

<sup>330</sup> See Sanbonmatsu et al., *supra* note 308, at 893.

do not have an incentive.<sup>331</sup> We obviously cannot pay jurors to project accurately, but it may be worthwhile to remind them that inaccurate mindreading will lead to a wrongful conviction or a false acquittal, both very costly outcomes. As I have discussed elsewhere, research indicates that the fear of a wrongful conviction primes jurors to consider alternatives to the prosecution's account of the defendant's guilt.<sup>332</sup> That same fear, if made salient, may also motivate jurors to do whatever they can to sufficiently adjust their simulations of the defendant's experience.

#### B. DEBIASING PROTOTYPING

Three techniques also exist for debiasing prototyping. First, although traditional jury instructions have no effect on the content of juror prototypes, a supplemental instruction that "attack[s] jurors' misconceptions about the target crime on a feature-by-feature basis, giving them specific information about how to revise the features contained in their naive representations," leads jurors to prototype far more accurately.<sup>333</sup> Smith, for example, presented mock jurors with the following supplementary kidnapping instruction, which was based on her earlier analysis of the disjunction between kidnapping prototypes and the legal definition of the crime:

Many people believe that kidnapping requires a ransom demand. However, a person can be found guilty of kidnapping even when ransom is *not* demanded, and even when the motive for the crime is not money. It is also not necessary that the victim of the kidnapping be a child. Adults can be victims of kidnapping as well. When the victim is a child, it is not necessary that the kidnapping result from a custody battle. A person can be found guilty of kidnapping even when that person is not involved in a custody battle. It is also not necessary that the victim of the kidnapping be taken to another location. A person can be found guilty of kidnapping even if the victim is not taken away . . . .<sup>334</sup>

Jurors who heard this instruction "no longer showed an effect of typicality, with equally high conviction rates for typical and atypical kidnappings," and more accurately determined when a kidnapping did not satisfy the legal requirements of the crime.<sup>335</sup>

Smith's research is extremely encouraging, because there is no reason jurors could not routinely be given similar supplemental instructions. There is, however, one difficulty worth mentioning: to maximize the debiasing effect of a supplemental instruction, a judge would have to determine not

---

<sup>331</sup> See Epley et al., *supra* note 68, at 332.

<sup>332</sup> See Heller, *supra* note 277, at 281-82.

<sup>333</sup> Smith, *supra* note 102, at 533.

<sup>334</sup> *Id.* at 529-30.

<sup>335</sup> *Id.* at 533.

only how the features of juror prototypes *typically* deviate from the formal legal elements of a crime or defense, but also—and far more importantly—how the prototypes *actually held by the jurors in the case* deviate from those elements. As we have seen, prototypes differ greatly between individuals. A general supplemental instruction would thus risk failing to revise the prototype features that are the most in need of revision.

The second and third debiasing techniques both concern personological prototypes: stereotypes of the “typical” battered woman or “typical” auto thief. Such stereotypes are particularly difficult to alter,<sup>336</sup> because they are “are deep seated, resistant to external pressures and strategic processes, and stable across time and situation.”<sup>337</sup> Nevertheless, two techniques have successfully moderated their influence on social perception. First, individuals can be asked to imagine someone who does not fit their stereotype. Studies have shown that a “consider the opposite” imagination task significantly reduces an individual’s stereotype expectancies when he later mindreads a specific target.<sup>338</sup> Second, individuals can be asked to take the target’s perspective prior to mindreading. Research indicates that such perspective taking “increase[s] the expression of positive evaluations of the target, reduce[s] the expression of stereotypic content, and prevent[s] the hyperaccessibility of the stereotype construct.”<sup>339</sup>

Both techniques could be utilized in the legal context. Voir dire could be structured to determine whether jurors possess inaccurate stereotypes (person-prototypes) of the defendant. If they do, they could be then asked to either engage in counterstereotype imagery or imagine how they would think and feel if they were the defendant before being sent to the jury room to deliberate. Either technique would reduce the likelihood—perhaps significantly—that the jurors would permit their stereotypes to undermine the accuracy of their mindreading.

The second technique, of course, essentially involves encouraging jurors to mindread through projection instead of prototyping. That is clearly a viable strategy when inaccurate person-prototyping seems

---

<sup>336</sup> See Irene V. Blair et al., *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 J. PERSONALITY & SOC. PSYCHOL. 828, 829 (2001).

<sup>337</sup> Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 PERSONALITY & SOC. PSYCHOL. REV. 242, 243 (2002).

<sup>338</sup> See *id.* at 249 (reviewing studies). It also reduces the likelihood that the individual will produce stereotype-consistent false memories—another salutary effect. See Lenton & Bryan, *supra* note 317, at 485.

<sup>339</sup> Adam D. Galinsky & Gordon B. Moskowitz, *Perspective-Taking: Decreasing Stereotype Expression, Stereotype Accessibility, and In-Group Favoritism*, 78 J. PERSONALITY & SOC. PSYCHOL. 708, 720 (2000).

particularly likely in a case—one in which the defendant is charged with a racially congruent crime, for example. In such situations, projection may be a better option than mindreading.

#### IX. CONCLUSION

*Actus non facit reum nisi mens sit rea*—“the act does not make a person guilty unless the mind is also guilty.” Such a simple maxim, yet one so difficult to apply. Aquinas did not believe that ordinary mortals could determine whether an individual possessed a guilty mind; to him, “God alone, the framer of the divine law, is able to judge the inward movement of wills . . . .” Legal scholars, by contrast, have always dismissed Aquinas’s skepticism, insisting—even taking as an article of faith—that jurors can, in fact, determine the mens rea of those whom the criminal law seeks to punish.

This Article has attempted to take that debate seriously, not to resolve it. The traditional explanation of mindreading is clearly inadequate: jurors mindread through projection and prototyping, not through functionalist theorizing. But that does not mean that Aquinas is correct and jurors are incapable of accurate mindreading. Accurate mindreading *is* sometimes possible: namely, when a juror uses projection to determine the mental state of a defendant who is very similar to himself or applies a legally precise prototype to determine the mental state of a defendant who is very different.<sup>340</sup> It is also possible, given the necessary will, to improve the accuracy of both mindreading techniques. Whether they can be made accurate enough to justify the vast cognitive demands contemporary criminal law imposes on jurors, however, very much remains to be seen.

---

<sup>340</sup> See, e.g., Ames & Iyengar, *supra* note 163, at 281 (“Good judges may be typical population members who recognize their similarity and project as a result, or atypical population members who recognize their dissimilarity and, accordingly, curb their projection.”).

