

CHAPTER 37

MENS REA IN MORAL JUDGMENT AND CRIMINAL LAW

CARLY GIFFIN AND TANIA LOMBROZO

37.1 INTRODUCTION

IMAGINE that Annette started a small fire in her backyard with the intention that it spread and burn her neighbour's house. She was successful, and her neighbour's house caught fire. Now imagine Blane, who started a fire of exactly the same size in his backyard. Unlike Annette, his intention was only to burn some brush. But the fire accidentally spread, and his neighbour's house caught fire. Although Annette and Blane performed the very same physical actions, and those actions generated the very same consequences, our intuitive moral judgments about them are likely to differ: Annette is more blameworthy and more deserving of punishment. This intuitive moral judgment is mirrored by the legal system's treatment of criminal liability, where Annette could be found guilty of arson and sentenced to between three and eight years in jail (California Penal Code, §451b), whereas Blane could be guilty of unlawfully causing a fire and sentenced to a maximum of six years in jail, or possibly receive only a fine (California Penal Code, §452b).¹

This example illustrates the influence of an actor's *mental states* on our intuitive and legal judgments. Within the law, '*mens rea*'—Latin for 'guilty mind'—is a necessary element for being found guilty of most crimes, and empirical work in psychology typically finds that moral evaluations are strongly influenced by such information (e.g. Mikhail, 2011). But is this influence warranted? If so, when and why ought we to take an actor's mental states into account in determining their blameworthiness, their moral responsibility, or the kind and amount of liability or punishment that they deserve?

¹ Blane's sentence would depend on further investigation of his actions, including his mental state. How reckless was he? What precautions—if any—did he take? Annette's sentence could well also be mitigated or aggravated by factors such as the amount of harm caused, but in any case her intent makes her crime and her punishment more severe than Blane's. Both Annette and Blane's actions might also subject them to civil liability, a discussion beyond the scope of this chapter.

In the present chapter, we take up the topic of *mens rea* in moral and legal evaluation. Our aim is not to definitively answer the normative questions that typically motivate moral philosophers, but instead to consider what the legal system and empirical research on people's intuitive moral and legal judgments can tell us about when and why mental states in fact influence moral judgments. In particular, we focus on what criminal law within the United States and experimental psychology tell us about the basic contrast between committing a moral or legal violation *knowingly* versus *unknowingly*. When is this distinction relevant, and why?

Section 37.2 explores the legal mental state requirement of *mens rea*, including its roots in pragmatic and moral considerations. Section 37.3 reviews what psychological findings can tell us about the role of mental states in moral evaluation, highlighting several cases in which the law and lay intuition are in agreement, as well as cases in which moral considerations appear to influence mental state evaluations in ways the law would not sanction. Section 37.4 concludes with brief remarks about the possible normative implications of this research.

37.2 THE ORIGINS OF MENS REA: BLOOD FEUDS TO CHURCH DOCTRINE TO ACADEMIC FEUDS

In contemporary law, *mens rea* is defined as the state of mind indicating the culpability which is required by statute as an element of a crime (Legal Information Institute 2016). This means that for most crimes, the statute specifies what mental state the defendant must have had to be convicted of the crime. For instance, in our introductory example, Annette could be found guilty of arson because she 'willfully and maliciously' caused her neighbour's house to burn (California Penal Code, §451). However, Blane could be found guilty of the lesser crime of 'unlawfully causing a fire' if it was determined that he acted not 'willfully and maliciously' but 'recklessly' (California Penal Code, §452).² In the legal system, arson is associated with greater punishment than is unlawfully causing a fire, regardless of the fact that each act could cause the same harm. Thus, a different mental state on the part of the defendant can change which crime she is charged with and the penalties she faces, precisely because she does or does not meet the specified *mens rea*.

Broadly, *mens rea* captures aspects of knowledge, belief, and intent that are interesting to scholars and, as the example above illustrates, consequential to defendants. Today, the majority of criminal statutes in the United States require some level of *mens rea*,³ from intent on the high end of the spectrum to negligence on the low end.⁴ However, this was not always the

² If he did not even act recklessly, the charge could be reduced still further.

³ Although, especially at the federal level, the number of crimes that do not require *mens rea* (so called 'strict liability crimes') is proliferating in the United States (U.S. House of Representatives 2013). We will discuss strict liability crimes in more detail later in the chapter.

⁴ It is worth noting that while the majority of crimes require *mens rea*, not all statutes explicitly include a *mens rea* provision. Supreme Court precedent, however, has made clear that the requirement of *mens rea* should be read into a statute unless the language of the statute or the legislative history make clear that *mens rea* was intentionally omitted (see *United States v. Morissette* 1952).



case historically. The fact that *mens rea* was not always an integral part of the criminal law raises an interesting question: how did it come to be a central part of criminal liability?

The English system of law, on which the American system is based, began with no consideration for mental states. Its purpose was to assure the swift compensation of the victim or victim's family to avert bloodshed (Gardner 1993; Perkins 1939; Sayre 1932). Prior to the institution of a (more) formal system of law, a wronged party would seek compensation in the form of vengeance against the wrongdoer, and sometimes against the wrongdoer's entire family. It was so-called 'blood feuds' that English law originally sought to avoid. In this context, it didn't matter whether someone killed a neighbour's sheep accidentally or on purpose; what mattered was that the sheep be replaced before any vengeance was sought. However, more complicated cases were difficult to fit into this model. For some actions, it was difficult to define compensation: how high should a monetary fine be for taking a human life? To whom should one pay a fine if the deceased had no family? The answers to such questions were not clear in the twelfth century, and they remain unclear in our own.

The law thus evolved to consider more than monetary fines, sometimes imposing sentences that included imprisonment. What the defendant knew or intended seemed more important when the emphasis shifted from compensating the victim to punishing the wrongdoer. To acknowledge mental states that might fall short of complete culpability in this new context, the king (for instance) might award pardons in cases of clear accident or self-defence (Gardner 1993). This shift was also supported by the church's rising influence in the twelfth and thirteenth centuries: prevailing Christian doctrine did not condone punishment without moral desert, and moral desert was lacking if a culpable mental state was not present. In fact, the first reference to *mens rea* as a consideration of English law is drawn from a canonical book (Sayre 1932; Hall 1960). As the church gained influence, the law responded to these concerns.

While scholars acknowledge the Judaeo-Christian roots of the *mens rea* doctrine, they disagree about whether *mens rea* should be tied to moral desert. On one side are those who maintain that the relevance of *mens rea* is indeed tied to moral desert, i.e. that acting knowingly in a particular case is wrong in large part due to the mental state involved, and that it is deserving of punishment for that reason. One famous defence of morality's position in the law argued that immoral acts tear at the very fabric of society: 'There are no theoretical limits to the power of the State to legislate against treason and sedition, and, likewise I think there can be no theoretical limits to legislation against immorality' (Devlin 1959). A less extreme formulation argues that both the law and morality are systematizations of what we owe each other, and that a *mens rea* assessment is not only an assessment of liability but also of wrongdoing and desert for violating community expectations (Morse 2004). On the other side are those who argue that conflating *mens rea* and moral desert does a disservice to the doctrine, allowing assessments of culpability to be unfairly influenced by extraneous factors, such as the defendant's character or the nature of the crime.⁵ They believe that assessing a defendant's *mens rea* should be strictly about the presence or absence of culpable mental states—whether

⁵ Another formulation of this argument might be to note that certain defences argue against moral condemnation even when *mens rea* is present. For instance, someone who kills another person in self-defence clearly did so intentionally, but their reason for doing so reduces or perhaps erases moral condemnation. This suggests that the mental states of an actor and the morality of their action can be separated.

the defendant performed the act *knowingly*, for instance—not about whether the defendant is a bad person who deserves to be punished (e.g. Miller 2001; Gardner 1993; Hart 1963).

Perspectives on the relationship between *mens rea* and moral desert should also be influenced by scholars' beliefs about why perpetrators ought to be punished in the first place. Those who advocate for punishment as *retribution* might argue that if an actor intended her actions to bring about a harmful consequence, she is more deserving of punishment. However, those who advocate for punishment on utilitarian grounds—such as deterrence, incapacitation, or rehabilitation—might instead consider *mens rea* only insofar as it relates to these bases for punishment (Craswell and Calfee 1986; Stahlkopf et al. 2010). Relevant questions could include: Is punishment more or less likely to function effectively as a deterrent when it applies to unintended consequences? Is incapacitating an actor more important when the harmful action was intended (perhaps because the perpetrator is more likely to cause harm again)? Is rehabilitation (in general or in some particular form) less likely to be necessary or effective when an action was unintended? The answers to these questions could determine whether and how advocates for a particular approach to punishment are inclined to treat *mens rea* in verdicts and sentencing.

37.3 THE PSYCHOLOGY OF *MENS REA*: DO MENTAL STATES MATTER?

Debates about the nature of *mens rea*, and about the moral commitments that underlie its application in the law, have proceeded largely independently of empirical research on human moral judgment. However, the kinds of questions discussed in §37.2 present natural analogues as research questions for psychology. How important are mental states when judging the actions of others? When do they matter more or less? Do people carve mental states into the same categories as the law? If not, what can the mismatches teach us? This section considers what insight psychology can bring to bear on these questions. Section 37.3.1 will review psychological research on the impact of mental states on intuitive moral judgments. Section 37.3.2 will examine work on whether laypeople can or do make the same mental state distinctions as the legal system. Finally, §37.3.3 will explore additional factors that influence mental state ascriptions.

37.3.1 (When) do mental states impact judgments?

In this section, we review empirical work that considers the role of mental states in people's intuitive moral judgments. In shifting from the law to human psychology, it is important to note that the use of terms such as 'knowingly' or 'intentionally' within a legal statute may not correspond to the use of these term within psychology or casual speech. Moreover, a psychologist studying the influence of mental states on moral judgment could be concerned with broader notions of knowledge, belief, intent, or some combination, without differentiating between these mental states or aiming to align them with corresponding legal usage. Nonetheless, psychology offers a rich set of findings concerning the role of mental states in



moral judgment, and we begin with this work before considering judgments within the legal domain.

Research overwhelmingly suggests that when assessing the actions of another person, that person's (inferred) mental state strongly influences our judgments.⁶ Justice Holmes famously noted that even a dog knows the difference between being kicked and being tripped over (1881), and humans are no different. Studies find that mental states are an important consideration when assessing wrongness, permissibility, blame, or punishment (e.g. Cushman 2008; Giffin and Lombrozo 2016; Mikhail 2009; Mueller, Solan, and Darley 2012; Young et al. 2011), with an especially pronounced role for the former two judgments (Cushman 2008). For example, someone who intentionally places lumber on a trail intending to cause bicyclists to crash is judged more blameworthy than someone who drops the lumber while transporting it to a building site (Shen et al. 2011). Similarly, studies have found that actors are judged more harshly for bringing about an outcome that they intended as opposed to one they merely foresaw (e.g. Murray and Lombrozo 2017; Mikhail 2009; Lagnado and Channon 2008). These and other studies have found, generally, that the greater the intent (from unknowingly to intentionally causing an outcome), the greater the moral condemnation, blame, and punishment (e.g. Mikhail 2009). One study even found that actors are blamed more for 'wicked desires' than for generally bad character (Inbar, Pizarro, and Cushman 2012).

Results such as these support the primacy of mental-state judgments when assessing criminal wrongdoing, but other research finds that the beliefs or intent of an actor are more important when judging some actions than others. Young and colleagues, for example, found that an unknowing harm violation (e.g. giving peanuts to someone you did not know has a peanut allergy) was judged less morally wrong than an unknowing purity violation (e.g. having sex with someone you did not know was your long-lost sibling), while a failed attempt to harm was judged more morally wrong than a failed attempt to commit incest (Young and Tsoi 2013; Young and Saxe 2011; see also Barrett et al. 2016). This suggests that for harm violations, intent is crucial, but for purity violations, the action itself—rather than the intent—plays a relatively greater role. The authors suggest that this difference could be due to the different functional roles of moral condemnation regarding harm versus purity. In the former case, we may want to predict and police an individual's future behaviour to evaluate (for example) the probability of causing future harm; for such an evaluation, mental states are important. Purity violations, by contrast, tend to be 'victimless' (Young and Saxe 2011). People may care less what an actor intended or knew if the consequences of that action affect only the actor herself. Moreover, if maintaining purity norms is important because it serves as a social signal of group membership, then a purity violation could compromise an individual's or group's status even if committed accidentally.

Another line of work suggests an additional factor that influences the relative importance of an actor's mental state in evaluating the violation of a rule: whether the rule is moral (such as prohibition on battery) versus conventional (such as a dress code; see Turiel 2008; Weston and Turiel 1980). Giffin and Lombrozo (2018) tested this idea by having people evaluate an actor who violated a moral or conventional rule, either knowingly (e.g. he knowingly

⁶ It is worth noting that the vast majority of decisions made by judges or jurors are made based on what they infer the defendant's state of mind to be. Even if the defendant explicitly states what she knew and intended, judges and jurors need not take them at their word.

injured another student, or knowingly wore a shirt that violated a dress code) or unknowingly (e.g. he mistakenly and innocently injured another student, or mistakenly and innocently violated the dress code). The moral rules were all intrinsically linked to harm, whereas the conventional rules were linked to wrongdoing only contingently: violations were wrong *because they violated the rule*. These studies found that whether the actor violated the rule knowingly versus unknowingly had significantly less impact on later judgments of moral wrongness and punishment if the rule that was violated was conventional, such as a dress code, rather than moral, such as hitting. Knowingly wearing the wrong shirt was not significantly worse than doing so unknowingly, but knowingly hitting another student was significantly worse than doing so accidentally.

Interestingly, a parallel distinction is found within the law. As stated at the beginning of the chapter, to be found guilty of most crimes in the United States, the prosecution must prove that the defendant had the requisite *mens rea*. Most crimes, but not all. ‘Strict liability’ crimes are the exception: these are crimes for which the prosecution is not required to prove that the defendant had any *mens rea*. Speeding is an example: a person can get a speeding ticket regardless of whether she knew she was speeding. Even if she could prove that her speedometer was broken, through no fault of her own, she could still be forced to pay her ticket. Strict liability crimes have no *mens rea* requirement, so her faultless ignorance is no defence.⁷

One justification for treating strict liability crimes so differently argues that strict liability crimes tend to be *malum prohibitum*—wrong as prohibited—while other crimes tend to be *malum in se*—wrong in themselves (*United States v. Morissette* 1952). That is, hitting is wrong whether or not it is illegal because it causes pain and distress, but driving 45 miles per hour on a particular street is wrong only *because* it is illegal. This distinction is nicely mirrored by the distinction scholars have made between conventional violations (which are wrong because prohibited) and moral violations (which are wrong in themselves), and scholarly descriptions only enforce the analogy. Justice Jackson, writing for the Supreme Court on strict liability crimes, said:

While such offenses do not threaten the security of the state in the manner of treason, they may be regarded as offenses against its authority, for their occurrence impairs the efficiency of controls deemed essential to the social order as presently constituted. In this respect, whatever the intent of the violator, the injury is the same, and the consequences are injurious or not according to fortuity. (*United States v. Morissette* 1952)

Regardless of rationale, legal scholars have long criticized strict liability in criminal law, arguing that a lack of knowledge is an important mitigating fact when judging a defendant. Based on the overwhelming evidence that intent matters in moral judgment, we might have predicted they were correct. On the other hand, the findings from Giffin and Lombrozo

⁷ Some strict liability crimes, in some states, do allow for a reasonable mistake-of-fact defence. For instance, in some states it is a defence to a charge of statutory rape that the defendant reasonably believed his partner was of age, for instance meeting them in a bar where they were being served alcohol. However, in *Garnett v. State*, it was no defence to statutory rape that the defendant had an IQ of 52, and a mental age of approximately 11 or 12, younger than the chronological age of his victim. The court ruled that in the state of Maryland the crime was strict liability, so it did not matter that Garnett was perhaps not even capable of having the requisite *mens rea* (1993).



(2018)—that acting knowingly versus unknowingly has a weaker impact for conventional violations than for moral violations—suggests that intuitive moral judgments could actually mirror the law’s treatment of strict liability crimes. Indeed, that’s what Giffin and Lombrozo (2016) found. As we describe below, participants judged ignorance to be significantly less mitigating for strict liability crimes than for crimes that required *mens rea*.

Across several studies, participants read vignettes describing an actor who committed a crime knowingly or unknowingly, where the crime was either strict liability or not strict liability (though this was not stated to participants). Participants then indicated how *wrong* the criminal act was, and how much punishment was warranted. For crimes that are not strict liability, the findings mirrored those from prior work: a knowing act (e.g. intentional theft) was judged more harshly than its unknowing counterpart (e.g. accidental theft). But for strict liability crimes, *mens rea* had little effect—an unknowing act (e.g. unintentionally speeding, or having sex with a mature-looking 16-year-old who was not known to be underage) was judged about as wrong as a knowing act (e.g. intentionally speeding, or knowingly having sex with a 16-year-old). This was true for a range of strict liability crimes, from speeding to shooting a migratory bird.

While these findings may be surprising to some legal scholars, they are much less surprising in light of the non-legal findings discussed above. Both conventional violations and strict liability crimes are (at least to some extent) *malum prohibitum*. The acts in themselves—wearing a particular-coloured shirt, or driving at 45 miles per hour—are not inherently wrong, but only become wrong in certain contexts because a rule (such as a dress code or a speed limit) prohibits them. While there might be good reasons for imposing such rules, their particulars are arbitrary in the sense that they could reasonably have been otherwise: the dress code could have specified a red shirt rather than blue; the speed limit could have been 55 rather than 45. As a result, intentionally violating a conventional rule is only contingently linked to an intention to harm—it is not clear that the perpetrator of an intentional violation intends any negative consequences beyond the rule-breaking itself. A person who intentionally drives over the speed limit, for example, does not (typically) intend to cause an accident, and if she did, she would be guilty of reckless driving, property crimes, vehicular assault, or even vehicular homicide (depending on the facts), in addition to speeding. Conversely, hitting someone causes pain and distress in addition to breaking a rule or law. An actor who intentionally hits typically⁸ intends these additional consequences, making an intentional violation of such a rule significantly worse. Even within the moral domain, where rule violations are typically linked to (potential) harm, participants who are told that an actor *believed* the rule had been set arbitrarily—such that violating it did not have an intrinsic consequence—found ignorance significantly less mitigating, just as they had for conventional violations and strict liability crimes (Giffin and Lombrozo 2018). This

⁸ While intentionally hitting typically brings with it the intention to bring about the harmful consequences of hitting, it’s possible to imagine circumstances in which an agent hits intentionally (e.g. to dislodge food from the throat of a choking victim), but does not intend any concomitant harm. We do not intend to assert a *necessary* link between intentionally breaking a moral rule and intentionally bringing about the associated harm, but rather to point to a difference in the inferences and evaluations that each type of action will typically license. In the context of this chapter, it is also important to note that someone who hit another person to dislodge a piece of food would not be considered by our legal system to have the same *mens rea* as someone who intentionally hit another person for less altruistic reasons.

was true even in cases where the actor was mistaken in his belief, and the violation *did* cause harm. Thus, even within the moral domain, an actor's mental state is more important to the extent that it suggests the actor intended to cause bad outcomes beyond the rule-breaking itself.

In sum, prior work investigating the influence of mental states on moral judgments reveals not only that mental states often matter, but also that the role of mental states is not uniform across types of mental states (e.g. intention versus foresight; e.g. Murray and Lombrozo 2017; Shen et al. 2011; Mikhail 2009; Lagnado and Channon 2008), types of judgments (e.g. wrongness versus punishment, Cushman, 2008), or types of transgressions (e.g. harm versus purity, Young and Tsoi 2013; Young and Saxe 2011; moral versus conventional, Giffin and Lombrozo 2018). These studies suggest that the mental state with which an actor acts is important to moral condemnation, blame, and punishment judgments to the extent that it signals something further about the actor, such as her additional beliefs, intentions, or likely future behaviour.

37.3.2 Do people parse mental states in the way the law expects?

The research in the previous section suggests that at a broad level, people and the law agree about the role of *mens rea* in moderating the severity of a crime. But do people make the same mental state distinctions that the law does? Are they even capable of doing so?

In some cases, folk judgments correspond well with the law. For instance, the difference between intentionally and knowingly committing an act has important consequences for moral judgments (Ames and Fiske 2015), and this distinction mirrors the Model Penal Code (MPC) distinction between purposely and knowingly.⁹ Research confirms that people can reliably differentiate between the *mens rea* categories of purposeful and knowing, as well as reckless and negligent, with or without jury instructions, indicating that these categories do track intuitive notions (Mikhail 2009; Shen et al. 2011; Mueller et al. 2012). However, not all statutes differentiate between purposely and knowingly (Mikhail 2009), suggesting that people may sometimes make finer distinctions than statutes do. The converse is also true: sometimes the law makes distinctions that laypeople do not reliably make themselves. In one study, participants performed at just above chance levels when asked to differentiate between knowing and reckless, and this conflation was also seen in their judgments concerning appropriate levels of punishment (Shen et al. 2011).

There are also cases in which legal and folk notions correspond roughly but imperfectly. For instance, people blame and punish an actor more if he acted intentionally, but research

⁹ The MPC is a document first drafted by the American Law Institute in 1962 to try to offer some clarity and uniformity to penal statutes, and the majority of states now base their codes, at least in part, on the MPC, making its definitions important starting points for discussion. The MPC defines conduct as performed 'purposefully' '(i) if the element involves the nature of his conduct or a result thereof, it is his conscious object to engage in conduct of that nature or to cause such a result; and (ii) if the element involves the attendant circumstances, he is aware of the existence of such circumstances or he believes or hopes that they exist' (MPC, §2.02(a)). The MPC defines conduct as performed 'knowingly' '(i) if the element involves the nature of his conduct or the attendant circumstances, he is aware that his conduct is of that nature or that such circumstances exist; and (ii) if the element involves a result of his conduct, he is aware that it is practically certain that his conduct will cause such a result' (MPC, §2.02(b)).



suggests that judgments of intentional action are themselves composed of multiple elements, including intention, skill, and awareness of fulfilling the intention while performing the action (Malle and Nelson 2003). The concept of intention, as used in *mens rea* statutes, involves a belief that an action will cause a result and the desire to bring that result about; skill is not explicitly identified as a component of the *mens rea* evaluation.

A final point worth making is that psychological models of moral evaluation don't necessarily map onto the legal process. The Path Theory of blame (Malle, Guglielmo, and Monroe 2014), for example, suggests that upon detecting an agent who caused an event, people assess blame by considering whether causing the event was intentional. If it was intentional, they consider reasons why the actor may have acted in this way. If it was unintentional, they consider whether the actor had the duty and ability to behave differently. This model works from outcome to mental states, rather than beginning with an evaluation of the mental states that preceded the action. This is counter to the way the law conceptualizes *mens rea*, insofar as a determination of the actor's *mens rea* should not depend on features of the outcome that could be incidental to the mental states that produced it. It is, however, much like what jurors are asked to do: consider an outcome and imagine the mental states that preceded it.

In sum, criminal law and the people it governs appear to be in broad agreement about when *mens rea* is important, and in the general idea of increasing moral culpability and punishment as defendants progress from negligently to purposefully performing an act. People do not, however, evaluate all and only those mental states specified within legal statutes, nor do they do so in a manner that follows the legal process. Perhaps part of the problem with mapping *mens rea* categories onto human psychology stems from the fact that the mechanisms underlying moral judgments are not necessarily accessible to us introspectively (e.g. Hauser et al. 2007). Legal concepts could involve an explicit articulation of concepts that laypeople apply implicitly, but when the explicit articulation and implicit understanding fail to correspond, people may not be aware of whether or why this is the case.

37.3.3 What factors influence *mens rea* ascriptions?

Beyond the factors discussed in §§37.3.1 and 37.3.2, research suggests that inferences concerning mental states are influenced by a variety of additional considerations, only some of which are condoned by the law. For instance, research shows that evaluations of an actor's character can change attributions of intent, and thereafter, assessments of blame and punishment. That is, actors who perform identical physical actions are sometimes judged to have different mental states and levels of culpability based on what kind of a person they seem to be, or what kinds of acts they are accused of committing (Nadelhoffer 2006). One study found that merely changing the description of the person committing the act, from a loving, doting aunt to a self-interested slacker, significantly increased participants' attributions of responsibility, causality, and blame for an identical action (Nadler and McDonnell 2012).

Research also finds that when people perceive an actor to have committed a negative, norm-violating behaviour, they perceive that person to have had a greater desire to cause that outcome than when an actor committed a positive, norm-violating act (Guglielmo and Malle 2010; see also Knobe 2003; Uttich and Lombrozo 2010). The moral valence of the act, then, changes mental state attributions. This could be particularly troubling for defendants charged with especially violent or heinous crimes. The charge itself could make it more likely

that jurors believe the defendant had the requisite *mens rea*. Even beliefs about a person that are generally advantageous, such as seeing a person as being of higher status, can negatively impact mental state attributions: actors who are perceived to have higher status, due either to privilege of birth or to achievement through work, were judged to be acting more intentionally than actors of low status who committed the same offence (Fragale et al. 2009).

Mens rea assessments could also vary based on the outcome a juror *wants* to realize, a phenomenon generally referred to as motivated reasoning. When engaged in motivated reasoning, a person who wants to punish an actor constructs an interpretation of facts that makes that punishment acceptable (Sood 2013; Sood and Darley 2012). For example, when participants were told that the defendant must have caused harm in order to be punished for an act, participants indicated that harm was caused, *but only if the defendant also espoused views with which the participant disagreed*. If the defendant espoused views similar to those of the participant, participants were significantly less likely to say that the defendant's act (of being nude in public) caused any harm (Sood and Darley 2012). The motivation of the participant has also been shown to directly impact *mens rea* assessments. When told that an employer must have acted intentionally to cause harm in order for an injured worker to collect more money, participants overwhelmingly said the employer had *intended* the worker harm. When later asked to classify the employer's knowledge, most correctly classified the behaviour as knowing, reckless, or negligent depending on the condition—not as intentional (Mueller et al. 2012). This shows it was not that the participants could not make the requisite distinctions; rather, they chose not to in order to attain the outcome they wanted.

To what extent are these influences on *mens rea* ascriptions inconsistent with the law? Legal decision-makers—judges and jurors—are not expected or asked to leave all their prior experience outside the courtroom. Instead, they are asked to make decisions based on the evidence presented *with reference* to their life experiences. For instance, when determining whether a witness is telling the truth, jurors are instructed to use their own experience. These same jury instructions, however, tell jurors to set aside bias and prejudice (CALCRIM 105).¹⁰ Jurors are expected to use their experiences, but they are also expected to ignore their biases. A juror who decides that a defendant who is described as habitually lazy and inattentive is more responsible for a bad outcome—because she likely didn't take care in this instance either—may be reasonably interpreting the evidence presented (i.e. Nadler and McDonnell 2012).¹¹ However, jurors who decide that harm has been caused merely because they do not like the views the defendant holds (Sood and Darley 2012) have clearly crossed a line the law did not intend. Even in the 'slacker' case, jurors who take the inference one step further and decide that the defendant is guilty not because he was careless (which may well be relevant), but because he is generally a bad person, are making a judgment that goes beyond that sanctioned by the law. The psychological findings thus lead us to think seriously about

¹⁰ The relevant jury instructions for California read, 'You alone must judge the credibility or believability of the witness. In deciding whether testimony is true and accurate, use your common sense and experience. You must judge the testimony of each witness by the same standards, setting aside any bias or prejudice you may have' (CALCRIM, 105, cited as Judicial Council of California Criminal Jury Instructions, 2018).

¹¹ This use of 'habit' evidence is expressly allowed by the Federal Rules of Evidence. 'Evidence of a person's habit or an organization's routine practice may be admitted to prove that on a particular occasion the person or organization acted in accordance with the habit or routine practice' (Fed. R. Evid. 406).



the functions of our moral evaluations and how they are best achieved, but also about the function of the legal system and how human psychology could contribute to or interfere with its operation.

37.4 NORMATIVE IMPLICATIONS

In cases where intuitive moral judgment conflicts with criminal law, a natural, normative question arises: how *ought* we to evaluate defendants? Should the law be adjusted to conform more closely to human psychology? Or are people simply making errors, where the law offers the correct normative benchmark against which to evaluate our more intuitive or reflective judgments?

The answers to these questions are of course complex, and our own suspicion is that both perspectives are partially correct: in some cases intuitive moral judgments may be capturing important normative dimensions (in which case legal reform should focus on better capturing human judgment); in other cases people are making consequential errors (and legal reform could focus on circumventing rather than accommodating the relevant aspects of human psychology). Moving forward, we offer two guiding suggestions.

First, when it comes to both legal and moral judgment, there may be value in adopting a functional perspective. We can ask: what roles do moral condemnation and punishment play in fostering prosocial behaviour and other desirable ends? How can benefits for individuals and communities best be realized? The answers to these questions have potential implications for the role of *mens rea* in moral and legal judgments. For instance, if an assessment of *mens rea* is made in the service of evaluating how likely a perpetrator is to reoffend, or how likely it is that a particular program of rehabilitation would be successful, then it may not be an error to rely on a variety of sources of information, including a person's character and the nature of the offence, as evidence rules encourage people to do. On the other hand, 'motivated' considerations, or those based on factors that could threaten fair treatment (such as race), are more appropriately considered errors in need of correction.¹² And, as already suggested, differences in evaluating harm violations, versus those related to purity or conventions, could also come from the different functions served by these evaluations. The sources of normativity here come in part (though only in part) from a functional perspective: focusing on the functions of particular judgments prompts us to ask how those functions are best realized, and to recognize that functions could vary across judgments, transgressions, and other features of a particular case.

Second, our chapter has discussed the legal system as a unitary entity, and 'human psychology' as a unitary entity. In fact, there is a great deal of important diversity to be acknowledged. Most of our discussion has focused on English and American criminal law, which not only differs from the legal systems found in other countries but also features its own heterogeneity. Within the United States, for example, states differ with respect to

¹² The law already acknowledges some of these potential dangers. At the federal level, the Federal Rules of Evidence limit the ability to introduce evidence of the defendant's character or prior bad acts (Fed. R. Evid. 404).

what *mens rea* terms are used in statutes. Some states use the term ‘intentionally’ in their statutes when the MPC, and other states, would use the word ‘purposely’, and it is debatable as to whether these two terms are describing identical or importantly different mental states.¹³ When it comes to the findings we cite from psychology, most research has recruited participants within the United States, many of them college students. Although some aspects of moral judgment appear to be quite robust across cultures (e.g. Hauser 2006), others vary across and within cultures (e.g. Shweder 1992; Haidt 2007). Even when it comes to mental-state evaluation, our present focus, claims may not be universal: a study of eight traditional small-scale societies found that the majority judged actions more ‘bad’ and assigned more punishment when intent was present (vs. absent); but the effect was not present in all cultures, and it varied in magnitude when it was present (Barrett et al. 2016). Future work should attend to such diversity both as a constraint on general claims and as a rich source of evidence. If a functional perspective is appropriate, one might expect such variation to track different roles for moral condemnation and punishment across societies and contexts.

37.5 CONCLUSION

Moral psychology and the law are increasingly in a symbiotic relationship. Moral psychology has exploded in recent years (e.g. Greene 2015), with psychologists studying an ever-wider range of questions about morality and mental states using a correspondingly increasing number of methods. Psychologists are realizing that the legal system offers a rich source of evidence concerning human moral judgment (see e.g. Mikhail 2011).

Legal systems require systematic attention to the relationships between knowledge, beliefs, intent, blame, moral condemnation, and punishment. Accordingly, the structures erected to sort and judge offenders can tell us something about how their architects understand these relationships. For instance, the empirical legal findings reviewed in this chapter used the structure of the legal system to learn new things about how people categorize mental states, and when they consider them to be informative. The legal system is a reflection of the minds and values of the people who constructed it. To the extent that we find mismatches between the legal system and laypeople—and certainly we do—the system provides an opportunity for anyone interested in morality to explore whether these mismatches show a failure of the system to accurately represent the people or a discrepancy between everyday judgment and our more reflective aspirations.

ACKNOWLEDGEMENTS

The authors would like to thank editors Manuel Vargas and John Doris for their help and guidance during the writing of this chapter, and John Mikhail for helpful comments and feedback.

¹³ New York, for instance, uses ‘intentionally’ instead of ‘purposely’ (N.Y. Penal Law §15.05).



REFERENCES

- Alicke, M. 1992. Culpable causation. *Journal of Personality and Social Psychology* 63(3): 368–78.
- 13, C.P.C., §§ 451 & 452.
- American Law Institute. 1962. *Model Penal Code*.
- Ames, D., and S. Fiske. 2015. Perceived intent motivates people to magnify observed harms. *Proceedings of the National Academy of Sciences* 112(12): 3599–3605.
- Ask, K., and A. Pina. 2011. On being angry and punitive: how anger alters perception of criminal intent. *Social Psychological and Personality Science* 2(5): 494–9.
- Barrett, H. C., A. Bolyanatz, A. N. Crittenden, et al. 2016. Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*: 113(17): 4688–93.
- Craswell, R., and J. Calfee. 1986. Deterrence and uncertain legal standards. *Journal of Law, Economics, and Organization* 2(2): 279–314.
- Cushman, F. 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108(2): 353–80.
- Cushman, F. 2015. Deconstructing intent to reconstruct morality. *Current Opinion in Psychology* 6: 97–103.
- Devlin, P. 1959. *The Enforcement of Morals*. Oxford: Oxford University Press.
- Fragale, A., B. Rosen, C. Xu, and I. Merideth. 2009. The higher they are, the harder they fall: the effects of wrongdoer status on observer punishment recommendations and intentionality attributions. *Organizational Behavior and Human Decision Processes* 108: 53–65.
- Gardner, M. R. 1993. The *mens rea* enigma: observations on the role of motive in the criminal law, past and present. *Utah Law Review* 635–750.
- Garnett v. United States*. 632 A.2d 797 (1993).
- Giffin, C., and T. Lombrozo. 2016. Wrong or merely prohibited: special treatment of strict liability in intuitive moral judgment. *Law and Human Behavior* 40(6): 707–20.
- Giffin, C., and T. Lombrozo. 2018. An actor's knowledge and intent are more important in evaluating moral transgressions than conventional transgressions. *Cognitive Science* 42(1): 105–33.
- Greene, J. 2015. The rise of moral cognition. *Cognition* 135: 39–42.
- Guglielmo, S., and B. Malle. 2010. Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin* 36(12): 1635–47.
- Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316: 998–1002.
- Hall, J. 1960. *General Principles of Criminal Law*. Indianapolis: Bobbs-Merrill.
- Hart, H. L. A. 1963. *Law, Liberty, and Morality*. Stanford, CA: Stanford University Press.
- Hauser, M. 2006. *Moral Minds*. New York: Ecco Press.
- Hauser, M., F. Cushman, L. Young, R. Jin, and J. Mikhail. 2007. A dissociation between moral judgments and justifications. *Mind and Language* 22(1): 1–21.
- Hindriks, F. 2010. Person as lawyer: how having a guilty mind explains attributions of intentional agency. *Behavioral and Brain Sciences* 33(4): 339–40.
- Holmes, O. W. 1881. *The Common Law*. Boston, MA: Little, Brown.
- U.S. House of Representatives. 2013. *Mens rea*: the need for a meaningful intent requirement in federal criminal law. House Report no. 113–46. Washington, DC: U.S. Government Printer.

- Inbar, Y., D. A. Pizarro, and F. Cushman. 2012. Benefiting from misfortune: when harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin* 38: 52–62.
- Judicial Council of California. 2018. *Criminal Jury Instructions*.
- Kahan, D. 2015. Laws of cognition and the cognition of law. *Cognition* 135: 56–60.
- Knobe, J. 2003. Intentional action and side-effects in ordinary language. *Analysis* 63: 190–94.
- Lagnado, D. A., and S. Channon. 2008. Judgments of cause and blame: the effects of intentionality and foreseeability. *Cognition* 108: 754–70.
- Legal Information Institute, Mens Rea. https://www.law.cornell.edu/wex/mens_rea, last accessed Oct. 13, 2021.
- Malle, B. F., S. Guglielmo, and A. E. Monroe. 2014. A theory of blame. *Psychological Inquiry* 25(2): 147–86.
- Malle, B. F., and S. Nelson. 2003. Judging *mens rea*: the tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law* 21: 563–80.
- Mikhail, J. 2009. Moral grammar and intuitive jurisprudence: a formal model of unconscious moral and legal knowledge. In *Psychology of Learning and Motivation*, ed. B. H. Ross. Cambridge, MA: Academic Press.
- Mikhail, J. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge, MA: Cambridge University Press.
- Miller, J. 2001. *Mens rea* quagmire: the conscience or consciousness of the criminal law? *Western State University Law Review* 29: 21–56.
- Morse, S. 2004. Inevitable *mens rea*. *Harvard Journal of Law and Public Policy* 27: 51–64.
- Mueller, P. J., Solan, and J. Darley. 2012. When does knowledge become intent? Perceiving the minds of wrongdoers. *Journal of Empirical Legal Studies* 9(4): 859–92.
- Murray, D., and T. Lombrozo. 2017. Effects of manipulation on attributions of causation, free will, and moral responsibility. *Cognitive Science* 41(2): 447–81.
- Nadelhoffer, T. 2006. Bad acts, blameworthy agents, and intentional actions: some problems for juror impartiality. *Philosophical Explorations* 9(2): 203–19.
- Nadler, J., and M. McDonnell. 2012. Moral character, motive, and the psychology of blame. *Cornell Law Review* 97: 255–304.
- Perkins, R. 1939. A rationale of *mens rea*. *Harvard Law Review* 52(6): 905–28.
- Sayre, F. 1932. *Mens rea*. *Harvard Law Review* 45(6): 974–1026.
- Schein, C., and K. Gray. 2017. The theory of dyadic morality: reinventing moral judgment by redefining harm. *Personality and Social Psychology Review* 22(1): 32–70.
- Shen, F., M. Hoffman, O. Jones, J. Greene, and R. Marois. 2011. Sorting guilty minds. *New York University Law Review* 86(5): 1306–60.
- Shweder, R. 1992. *Thinking Through Cultures: Expeditions in Cultural Psychology*. Cambridge, MA: Harvard University Press.
- Sood, A. 2013. Motivated cognition in legal judgments: an analytic review. *Annual Review of Law and Social Science* 9: 307–25.
- Sood, A., and J. Darley. 2012. The plasticity of harm in the service of criminalization goals. *California Law Review* 100(5): 1313–58.
- Stahlkopf, C., M. Males, and D. Macallair. 2010. Testing incapacitation theory: youth crime and incarceration in California. *Crime and Delinquency* 56(2): 253–68.
- Turiel, E. 2008. Thought about actions in social domains: morality, social conventions, and social interactions. *Cognitive Development* 23: 136–54.
- United States v. Morissette*, 342 U.S. 246 (1952).

- Uttich, K., and T. Lombrozo. 2010. Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition* 116(1): 87–100.
- Weston, D., and E. Turiel. 1980. Act–rule relations: children’s concepts of social rules. *Developmental Psychology* 16(5): 417–24.
- Young, L., and R. Saxe. 2011. When ignorance is no excuse: different roles for intent across moral domains. *Cognition* 120: 202–14.
- Young, L., and L. Tsoi. 2013. When mental states matter, when they don’t, and what that means for morality. *Social and Personality Psychology Compass* 7: 585–604.