Mental states: When they matter to moral judgments and when they don't

Tom R. Kupfer, Daniel M.T. Fessler, Juan Pablo Arroyave Hurtado, Christopher S. Kovel, Alex Lambert, Carlton Patrick, Harry Purser, and Paul Troop

Abstract

Mental states are central to moral judgment, yet the weight they receive is not uniform. We review evidence that mind-contingency varies systematically across (a) judgment types - wrongness, blame, punishment; (b) developmental stage and neurodiversity (in infancy or with autism); (c) cognitive conditions (time pressure, cognitive load); (d) social relationships (kin, strangers, out-groups); (e) legal regimes (*mens rea* vs. strict liability); (f) cultural ecologies (relational mobility, collectivism, "opacity of mind"); and (g) moral content (harm, purity, loyalty, etc.).

We integrate these findings with a cost-benefit model of mental-state deployment. Mental states are used facultatively - people seek out and integrate (or ignore) information about a wrongdoer's intentions when the expected benefits (accuracy, future coordination) outweigh the costs (cognitive effort, social risks). When integrated, that information is weighed flexibly, for example, unintentionality cues carry more weight for a relative than for a stranger. Because these trade-offs vary by context, moral judgments, and the norms and laws they shape - and are later informed by - span a continuum from fully mind-contingent to effectively mind-blind.

Citation:

Kupfer, T.R., Fessler, D.M.T., Arroyave, J.P.., Kovel, C.S., Lambert, A., Patrick, C., Purser, H., and Troop, P. (2026). Mental states: When they matter to moral judgments and when they don't. In *Research Handbook on Social Judgment,* J.F. Landy ed., pp. 173–192. Cheltenham, Gloucestershire: Edward Elgar Publishing.

Moral judgments permeate daily life, guiding whom we help, punish, or praise in every society. This chapter focuses on one aspect of moral cognition: inferences about a wrongdoer's mental states - intentions, beliefs, and desires - and the conditions under which those inferences shape moral judgment. Although moral evaluation also depends on the recipient's mental states, such as their consent, capacity for suffering, and perceived deservingness (Russell & Piazza, 2015; Gray & Wegner, 2009; Niemi & Young, 2016), this chapter concentrates on the agent's mind, which is the most direct and well-studied route through which mental-state inference shapes moral judgment. First, we describe how reliance on mental-state information differs across judgment types (wrongness, blame, punishment, praise, character) and how that reliance changes over development and under cognitive constraints. Next, we examine the impact of group membership and relationship closeness. We then consider how legal systems codify mental-state sensitivity through liability standards. Finally, we consider cultural and ecological pressures that expand or shrink the role of mental states in moral judgment, before offering an evolutionary account of when and why mental states matter.

People differ not only in whether they attend to a wrongdoer's intentions and beliefs, but also in both how strongly those considerations influence judgment, and which cues they prioritize. Across contexts, these differences reveal that the role of mental states in moral judgment is variable rather than fixed. In the final section, we suggest that this variability reflects an adaptive system that deploys and weights mental-state information according to the costs and benefits of the situation - a system that helps explain when moral judgment becomes more mind-contingent or more mind-blind.

Mental States Across Different Types of Moral Judgments

Moral judgments vary in the extent to which they rely on mental-state inferences, such as an agent's intentions, beliefs, and desires. Different types of judgment - wrongness, blame, punishment, praise, and character evaluations - assign varying weights to mental states versus the outcomes of actions. Moreover, outcomes may also shape mental-state evaluations and subsequent moral judgments through outcome-to-mind "backwards inference."

Wrongness, Blame, and Punishment

When judging how wrong an act is, people privilege mental states. In Cushman's (2008) experiments, an action was judged wrong chiefly to the extent that the agent desired and believed harm would occur, regardless of the harm's actual occurrence or magnitude. In contrast, blame and punishment judgments incorporate outcomes far more heavily. Cushman

(2008) and Alicke (2000) showed that identical acts of recklessness drew harsher blame and punishment when the act produced severe rather than minor harm. Kneer and Machery (2019) traced this severity effect to two sources: observers inferred greater negligence or carelessness for severe harms (an indirect path) and imposed stiffer penalties even after controlling for perceived negligence (a direct path). Outcome valence also matters: in Knobe's "side-effect" studies (2003a, 2003b), environmentally harmful side effects were judged intentional, whereas helpful side effects were not.

Together, these findings show that outcomes modulate how mental-state information is sought and weighted, especially for blame and punishment, whereas wrongness judgments depend largely on mental states.

Praise Judgments

Compared to blame judgments, praise judgments typically hinge less on intent. Whereas blame is sharply reduced when harmful outcomes are seen as uncontrollable, praise is not similarly reduced when a helpful outcome is beyond the individual's control (Pizarro et al., 2003). Likewise, perceived intent increases blame for negative behaviors more strongly than it increases praise for positive behaviors (Ohtsubo, 2007). However, motives do influence praise (Anderson et al., 2020). Actions perceived as driven by altruism or strong internal desire to help are judged more praiseworthy than the same actions done for self-serving reasons or under external pressure (Newman & Cain, 2014). These patterns may owe to the differing social costs associated with praise and blame. Blame and punishment impose costs on targets and can provoke retaliation (Balafoutas et al., 2014), whereas praise benefits targets and carries little risk for those bestowing praise (Anderson et al., 2020). Hence, calculations of blame and punishment involve a more stringent consideration of mental-state evidence, whereas praise is less stringent and hinges more on actions and outcomes.

Character Judgments

Character judgments rely heavily on inferences about an agent's mental states - such as their intentions, desires, and motives, consistent with the function of predicting behavior across contexts rather than evaluating a single act in isolation (Reeder, 2004; Uhlmann et al., 2015). Act-person dissociations in moral judgment show that people can approve of an act yet disapprove of the actor's character, and vice versa. For example, participants evaluated the character of someone who desired to kill one person to save others negatively, even when they judged the action itself to be morally right (Uhlmann et al., 2013). Similarly, a CEO who requests a marble table engraved with their likeness is judged as having worse moral

character than one requesting a monetary bonus, despite neither act being judged wrong (Tannenbaum et al., 2011).

Character evaluations can also shape how mental states are perceived, thereby influencing other moral judgments (Alicke & Zell, 2009; Nadler & McDonnell, 2011; Schwartz et al., 2022). In Alicke's (1992) research, despite the act and outcome being the same, more negative mental states (e.g., recklessness or selfishness) were attributed to a driver who caused an accident while speeding to hide cocaine than were attributed to a driver speeding to hide an anniversary gift. Similarly, Kliemann et al. (2008) found that selfish individuals were judged to have acted more intentionally than unselfish individuals despite performing the same action. These findings suggest that prior beliefs about character influence how mental states are inferred, and, in turn, this can shape character judgments.

Processes Linking Mental States and Outcomes Across Judgment Types

Several explanations have been proposed for why mental-state contingency varies across judgment types. According to motivated-blame or outcome-salience accounts, because harmful outcomes are emotionally engaging, they lead observers toward condemnation and, in the process, to revise the mental states they attribute, for example, upgrading perceived intent or recklessness. This outcome-to-mind, or backwards-inference, route changes judgment indirectly by altering the inferred mental-state evidence itself (Alicke, 1992, 2000; Buon et al., 2016; Patil et al., 2017). However, such outcome-driven accounts cannot easily explain why blame and punishment are influenced by harmful outcomes, but wrongness is not.

Cushman (2008) and colleagues (Martin & Cushman, 2016) proposed a two-process model of moral judgment. Wrongness judgments are thought to be influenced by a cognitive process sensitive to mental states, particularly intentions and beliefs. In contrast, blame and punishment judgments are influenced by both this and a second process that is sensitive to the causes of outcomes. Unintentional harms are judged not wrong because there is no culpable mental state, yet they are considered blameworthy and punishable because the agent caused a harmful outcome. Conversely, an intentional but failed attempt to harm is condemned as being just as wrong as a successful harm, even though no damage occurred, because the culpable intent is fully present (Young & Saxe, 2011).

More recent perspectives have argued that the role of negligence in explaining moral judgments has been overlooked (Monroe & Malle, 2019; Nobes et al., 2023). Judgments of unintended harm are associated with perceived negligence, including how probable a bad outcome was and whether appropriate care was taken (Kneer & Machery, 2019; Margoni et

al., 2023; Nelson-Le Gall, 1985; Nobes & Martin, 2022). The influence of outcome on judgments of unintended harm may therefore be due to inferences that the wrongdoer could and should have been more careful (Malle et al., 2014; Sarin & Cushman, 2024). Nonetheless, bad outcomes can still have a direct effect on moral judgments, unmediated by perceived negligence or other mental states, especially for punishment judgments (Kneer & Machery, 2019; Nobes et al., 2023).

Summary and Implications

In summary, the role of mental states varies across moral judgments. Wrongness judgments prioritize mental states such as intentions and beliefs, while blame and punishment judgments integrate both mental states and outcomes, with outcomes playing a particularly important role. Praise judgments, though less dependent on mental states, emphasize the motivational quality behind prosocial actions. Character judgments rely heavily on mental states to infer broader moral dispositions, often influencing other moral evaluations.

Together, these findings indicate that moral judgment is not uniformly mind-contingent, but rather is contextually calibrated. The weighting of mental states appears to depend on the *function* of the judgment: wrongness judgments diagnose moral intent, punishment judgments deter future harms, and praise and character judgments regulate relationships and reputation. Such variation suggests that the use of mental-state information is facultative - engaged more strongly when understanding an agent's mind advances the underlying social goal of the judgment and downplayed when outcomes or deterrence take precedence. We return to this idea in the final section.

Development of Mental-State Use in Moral Judgments

The moral judgments of young children, particularly those around age three, tend to focus heavily on outcomes rather than intentions. Children at this age often blame accidental and intentional harms equally (Piaget, 1932; Zelazo et al., 1996). As children grow, they increasingly incorporate mental-state information, marking a shift from outcome-based to intent-based judgments (Killen et al., 2011; Nobes et al., 2009). This transition is not uniform across judgment types.

By ages four to five, children begin to distinguish intentional from accidental harm in assessments of wrongness, becoming less likely to label accidental harms as "naughty" (Cushman et al., 2013; Margoni & Surian, 2017). However, development is uneven across judgments: while wrongness begins to integrate mental states earlier, punishment remains tied to outcomes until around age seven (Cushman et al., 2013; Margoni & Surian, 2017).

This pattern mirrors adult judgments, in which mental states are more pivotal for assessing wrongness and character than for determining blame or punishment, with some punishment still assigned to accidental outcomes (Cushman, 2008).

The increasing influence of mental states on judgments in older children correlates with improvements in mindreading abilities (Killen et al., 2011; Sodian et al., 2020). Children typically pass false-belief tasks around ages four to five (Wellman, Cross, & Watson, 2001). However, even younger children can pass implicit versions of these tasks (Scott & Baillargeon, 2017) and make intent-based moral judgments when mental-state cues are highly salient (Hamlin, 2013; Nobes et al., 2017; Woo et al., 2017). Therefore, the capacity to infer mental states may emerge early, but continues to develop alongside language and executive functions (Rakoczy, 2022). Moreover, young children often make outcome-based moral judgments despite understanding intentions (Nobes et al., 2009; Zelazo et al., 1996). This suggests a difference in *priority* rather than *ability*: children may emphasize that people should take care to avoid harm, hence they treat unintended harm as negligence. Correspondingly, children judge accidental harms more harshly when they result from lack of care than when care was taken (Nobes et al., 2009; Woo et al., 2017).

Negligent mental states continue to matter in adults (Nobes et al., 2017), which may help explain why adults assign some punishment to accidental harms (Nobes et al., 2023). In late adulthood (around ages 60–90), judgments again become more influenced by outcomes, with increased condemnation of accidental harms (Margoni et al., 2021). This shift is accompanied by stronger inferences of negligence (Margoni et al., 2019). Rather than mental states not mattering in early childhood or late adulthood, *different* mental states matter: carelessness looms larger than intent per se.

Autism and Neurodiversity

Difficulties with theory of mind are characteristic of autism (Baron-Cohen, 1995). Autistic individuals often rely more on observable outcomes in social reasoning (Frith & Happé, 1994), including in moral judgment (Dempsey et al., 2020). Compared with neurotypical individuals, autistic participants are more likely to judge accidental harm as equally punishable to intentional harm (Zalla et al., 2011). Some findings show that autistic individuals who understand faux pas and false beliefs nonetheless make outcome-focused moral judgments (Moran et al., 2010). However, when mental-state information is made highly salient, autistic individuals can make intent-based judgments similar to those of neurotypical individuals (Margoni et al., 2019b), suggesting that salience enhances the processing and prioritizing of mental-state information.

The reduced emphasis on intent may reflect difficulty processing subtle mental-state cues or integrating them with outcome information to arrive at moral judgments (Moran et al., 2010), especially for accidental harms where emotional arousal is high (Fittipaldi et al., 2023; Patil et al., 2017). Alternatively, as with young children and older adults, outcome-focused judgments may reflect priorities rather than inability: people should take care not to cause harm. That said, evidence on negligence judgments in autism remains limited.

Cognitive Constraints on Mental-State Use

Even for neurotypical individuals, mental-state inference can be cognitively demanding (Apperly, 2011). Under cognitive load, distraction, multitasking, or time pressure, people tend to default to observable outcomes rather than inferred mental states (Gilbert, Pelham, & Krull, 1998). In moral judgment tasks, cognitive load reduces consideration of intentions behind harmful actions (Buon et al., 2013; Martin et al., 2021). Notably, participants under cognitive load can still recognize whether harmful outcomes were intentional or accidental, yet they judge accidental harms more harshly - a pattern consistent with reduced deployment and weighting of intent when resources are limited. Load influences judgments of accidental but not attempted harm (Martin et al., 2021), plausibly because overriding a harm response in the face of conflicting innocent intent is more resource-intensive. Individual-difference findings converge: higher reasoning ability and a more deliberative (rather than intuitive) style predict forgiving accidental harms (Patil & Trémolière, 2021; Schwartz et al., 2022), either because better reasoners more readily identify innocent intentions, or because they more effectively integrate that information with outcomes.

Two important caveats apply here. First, most experiments use simplified, easily interpreted stimuli (verbal vignettes, cartoons, puppets). Real-life wrongdoings are often subtle and ambiguous, especially with respect to mental-state cues such as nonverbal timing and contextual expression. Such cases likely demand more cognitive resources than do typical experimental tasks, hence, outside the lab, capacity limitations may reduce the influence of mental states even further, particularly in autistic individuals. Second, harmful outcomes are not always obvious or emotionally salient, and may arise from complex or unforeseeable chains of events. Processing such outcomes can itself be cognitively costly, consistent with links between greater reasoning ability or cognitive resources and more consequentialist decisions in moral dilemmas (Patil et al., 2021).

Summary and Implications

In sum, mental-state inference is cognitively costly, time-consuming, and uncertain. When its costs are high or capacity is low, people deploy it less and weight it less, adjusting their standards for condemning or excusing accordingly. These shifts help explain why young children, autistic individuals, older adults, and adults under cognitive constraints often rely more on outcomes - and why, when intent evidence is sparse or costly, attributions of care or carelessness loom larger, since such cues are easier to infer from observable precautions and foreseeability, and are often codified in rules and protocols (see law section).

Together, these findings suggest that mental-state inference in moral judgment operates facultatively: it is engaged when the expected benefits of accuracy or coordination outweigh the cognitive or social costs, and flexibly weighted once deployed. Across development, neurodiversity, and cognitive load, we see the same principle: when mentalizing is difficult, uncertain, or inefficient, moral judgment shifts toward outcome-based evaluation. The same cost-benefit logic extends beyond the variables considered here, influencing how social and ecological contexts determine when mental states are considered and how much they matter.

Social Relationships and Group Membership

Most research in moral psychology has examined judgments of unrelated or anonymous strangers. Yet moral judgments, especially punishment, can deter only by imposing costs on wrongdoers. A functional account should therefore expect sensitivity to who the wrongdoer and victim are, with harsher judgments of enemies and strangers than of kin, allies, or exchange partners (Lieberman & Linke, 2007). These identities should also shape attention to and interpretation of mental-state information - that is, their deployment and weighting. Although people can aspire to impartiality (Baumard et al., 2013), several findings align with these functional expectations.

Ingroups and Outgroups

People exhibit ingroup favoritism in moral judgment: selfish or harmful acts are judged less immoral when performed by an ingroup member than an outgroup member (McKee et al., 2024; Valdesolo & DeSteno, 2007). Immoral behavior by ingroup leaders is punished less, especially when it serves group interests (Abrams et al., 2013). The degree of ingroup favoritism covaries with commitment and protectiveness toward one's group (Bocian et al., 2021).

Group membership also influences mental-state attribution. Both children and adults attribute mental states more quickly and accurately to ingroup than to outgroup members

(Adams et al., 2010; McLoughlin & Over, 2017), particularly when the outgroup is disliked or threatening (Ekerim-Akbulut et al., 2020). People may also ascribe more favorable motives to ingroup actors: ingroup-to-outgroup aggression is seen as driven by love of the ingroup, whereas outgroup-to-ingroup aggression is seen as driven by hate (Waytz et al., 2014). Ingroup members are also viewed as more capable of experiencing complex emotions (Waytz et al., 2010; Gaunt et al., 2002).

Consistent with these findings, identical mental-state information is interpreted less charitably for outgroups. In a U.S. sample, strong partisans discounted exculpatory information for outgroup wrongdoers: learning that a violation was accidental or intentional for good reasons (e.g., self defense) mitigated blame less for outgroup than for ingroup actors (Monroe & Malle, 2019, Study 6). Similarly, when judging political or racial outgroup wrongdoers versus ingroup wrongdoers, participants attributed more intention and more blame to outgroup members despite identical information (Staples & Plaks, 2024). In our terms, for ingroups, observers tend to deploy mental-state inference more, weight exculpatory cues more, and apply higher standards to condemn; for outgroups, the pattern reverses.

Relationships and Relational Value

People judge the same wrongdoing by relatives or friends as less wrong, less unacceptable, and less deserving of punishment than when the agent is a distant other or a stranger (Forbes & Stellar, 2021; Lee & Holyoak, 2020). People also judge wrongs that harm a relative more harshly than those that harm a stranger (Lieberman & Linke, 2007). Moreover, people endorse partiality: they view it as morally right that others are more lenient toward kin and friends than toward strangers (Hughes et al., 2016; Soter et al., 2021). However, when impartiality is made salient or normatively required, for example, a professor choosing between related and unrelated students, favoritism is judged unacceptable (McManus et al., 2020). Notably, these findings come largely from Western settings with strong impartiality values (Schulz et al., 2019), so partiality may be stronger in cultures emphasizing loyalty values.

Few studies directly test whether the relationship to the agent or victim changes how a wrongdoer's mental states are perceived, but the pattern above suggests it likely does. When a relative, compared with a stranger, commits a wrongdoing, observers may be more inclined to seek mental-state information (deployment), to emphasize exculpatory cues such as lack of intent or noble motives (weighting), and to raise the evidentiary standard for condemnation. This partiality should be most visible when intent and motive are ambiguous, as is common in real-world cases.

Summary and Implications

The degree and manner in which mental states matter to moral judgment depends on the identity of both wrongdoer and victim. Identical mental-state information is interpreted more favorably for ingroup versus outgroup wrongdoers, shifting judgments accordingly. The consistent finding that strangers are judged more harshly than relatives or close others suggests that mental-state attributions are similarly partial, though this specific link awaits direct tests. Other identity features (e.g., being both an ally and high-status) may combine to tilt how mental states are sought and weighted, so that the summed social value of the parties shapes how mental-state information is attended to and used.

These patterns show that moral cognition is facultative not only with respect to cognitive capacity but also social context. The perceived value of the wrongdoer and victim shapes both whether mental states are considered and how they are interpreted, consistent with moral judgment functioning to manage relationships and maintain alliances. Such shifts reflect the same cost–benefit logic described earlier: when maintaining a relationship or signalling loyalty outweighs the value of impartial accuracy, moral judgment becomes more lenient and more mind-contingent.

Mental State Requirements in the Law

Research on large-scale societies demonstrates a close link between moral intuitions and legal norms, suggesting that laws may be a cultural elaboration of our moral intuitions (e.g., Jones, 2001; Patrick, 2023; Sznycer & Patrick, 2020). In Anglo-American law, there are striking similarities between laypeople's inferences about a wrongdoer's motivations and the exculpatory factors used in actual sentencing decisions (Cushman, 2014; Mikhail, 2009, 2011), and intuitions about the motivational properties of anger map onto the legal doctrine of provocation (Patrick & Lieberman, 2017). In short, legal rules appear to reflect not only broad features of human psychology but also specific intuitions about mental states and wrongdoing.

Just as people routinely incorporate intent into culpability and punishment judgments, so too does the law (Mikhail, 2009; Robinson & Darley, 1995). In criminal cases, the fact-finding judge or jury is typically required to infer the mental state (or *mens rea*) of the accused, and punishment severity tracks mental state: intentional wrongdoing is punished more than recklessness; recklessness more than negligence; and accidents are usually not punished (American Law Institute, 1962; Robinson & Darley, 1995; Shen et al., 2011).

Nevertheless, there is variation. Even in criminal law, some offenses dispense with the requirement to prove a culpable mental state. Many jurisdictions recognize "strict" or "absolute" liability, where guilt can be established with little or no *mens rea* requirement (Cappelletti, 2022, p. 67; Child & Ormerod, 2023, pp. 93–94). Such rules often arise when the victim is vulnerable or the behavior is particularly dangerous. For example, in English law, causing a child to watch a sexual act permits a reasonable-belief defense about age only if the child was over 13 (s. 12, Sexual Offenses Act 2003). English law also recognizes gross-negligence manslaughter, which uses an objective "reasonable person" test, meaning that the fact-finder need not establish what the accused actually foresaw (Child & Ormerod, 2023, p. 123; see also Holmes, 1881, p. 108; and Tettenborn, 2023, paras 1–80). Similarly, perpetrators may be convicted where the absence of intention is self-induced - for instance, where a man killed his partner after taking LSD and believing he was fighting snakes (*R v Lipman* [1970] 1 QB 152).

The presence or absence of a mental-state requirement also varies with stakes. Where consequences for the accused are severe (e.g., long imprisonment, stigmatizing records), mental-state requirements are more common and standards to condemn are higher. In contrast, strict liability is common for quasi-criminal offenses, such as driving while intoxicated in England (s. 5, Road Traffic Act 1988). In the United States, strict-liability crimes often involve age or alcohol (e.g., driving under the influence, selling alcohol to a minor, statutory rape), though strict-liability elements also appear in more serious crimes, such as felony murder (Binder, 2008).

Civil law provides a further contrast. In contract law, breaches are generally treated as strict liability, where the breaching party is liable regardless of intent (e.g., *Cutter v. Powell* (1795); *Restatement (Second) of Contracts* (1981)). In tort, most claims are for negligence, which (like the crime of gross-negligence manslaughter) generally imposes an objective standard—what a reasonable person would have done, not what the actual person did (*Donoghue v Stephenson* [1932] AC 562; Cappelletti, 2022, p. 38). As a result, a novice learner may be assessed by the standards of a qualified driver (*Nettleship v Weston* [1971] QB 691).

Other areas of tort move further toward strict liability, especially for abnormally dangerous activities (*Rylands v Fletcher* (1868) LR 3 HL 330; *Animals Act* 1971; *Restatement (Second) of Torts* § 522; French Civil Code Arts. 1384–1386) and for novel, diffuse risks such as nuclear power, aviation, and space exploration (*Nuclear Installations Act* 1965; *Convention on Damage Caused by Foreign Aircraft to Third Parties on the Surface*

1952; *Convention on International Liability for Damage Caused by Space Objects* 1972). Similar approaches have been taken to activities with the potential to affect large numbers of people—for example, product liability (EU Council Directive 85/374; *Consumer Protection Act* 1987).

These rules also change over time (Tettenborn, 2023, paras 1–88). Historically, many U.S. and U.K. jurisdictions followed a "one-bite-free" approach, under which dog owners were not liable the first time their dog bit someone, on the assumption that they had no prior knowledge of the dog's dangerousness. Many jurisdictions have since shifted to strict-liability standards that hold owners responsible even for the first bite (Epstein, 2006; e.g., *Animals Act* 1971).

Lawmakers and scholars offer myriad rationales for strict-liability laws. A recurrent theme is shifting an additional burden of care to those best placed to avoid risk (serving alcohol, product manufacturing, hazardous activities, pet ownership). Other rationales include evidentiary difficulties in proving mental state, victim protection, curbing unjustifiable risk, and allocating risk to the party best positioned to prevent it (Cappelletti, 2022; Flume, 2022; White, 2003). For instance, some strict-liability offenses involving building codes or traffic violations have been justified by the administrative costs that would be incurred in gathering mental-state evidence, and the relatively low stakes of the fines assessed as penalties (Larkin, 2014).

There is no single accepted principle dictating when laws should include a mental-state requirement. Justifications vary across offenses and jurisdictions, and systematic quantitative mapping of mental-state requirements across legal systems - especially beyond the Anglo-American world - remains limited.

Nevertheless, several qualitative parallels mirror the psychological patterns reviewed elsewhere in this chapter. When inferring mental states is difficult (e.g., under cognitive load, time pressure, or developmental limits), moral judgments rely on them less (e.g., Martin et al., 2021). Similarly, laws are less likely to include a mental-state requirement when such proof is hard to establish in court. If a person creates a highly dangerous environment (e.g., keeping tigers on their property), law often dispenses with *mens rea* and focuses on outcomes and due care; likewise, moral judgments tend to rely less on exculpatory intent when outcomes are especially harmful (Cushman, 2008). Such parallels could reflect psychological intuitions shaping legal institutions over time, shared constraints on both domains (information, error, and deterrence costs), or both. Either way, they support the view that the role of mental states follows systematic, functional patterns at both the individual and

institutional levels - mapping, in our terms, onto shifts in facultative deployment (how much mental-state evidence is sought) and flexible weighting (how much it counts relative to outcomes and to care or foreseeability cues). Legal codes thus provide a formal mirror of the adaptive architecture underlying human moral judgment. Just as legal codes formalize context-dependent trade-offs in attending to mental states, cultural norms and ecologies shape those trade-offs in everyday moral judgment - a topic we examine next.

Cultural Variation in Mental States and Moral Judgment

An active question is whether cultures differ in how strongly moral judgments hinge on a wrongdoer's mental state (Barrett & Saxe, 2021; Park et al., 2022). Anthropologists have long noted that some Pacific societies emphasize the *opacity of mind* - the idea that one cannot truly know others' thoughts (Robbins & Rumsey, 2008). Ethnographic accounts of dispute resolution in such settings report attention to actions and consequences rather than intentions (Rumsey, 2013). Quantitative data likewise suggest that where social life is structured by dense kinship obligations or strong collectivist orientations, judgments rely more on outcomes than on mental states (McNamara et al., 2019; Curtin et al., 2020). Mental-state inference is not absent, but its deployment is less central, consistent with environments that make outcomes and social roles more informative than inner motives.

Kinship Intensity and Relational Mobility

One function of mindreading is prediction: inferring intentions to anticipate future behavior (Turpin et al., 2021). The value of such inference depends on *relational mobility* - how much choice people have over social partners (Thomson et al., 2018). When interactions are largely fixed by kinship or obligation, partner choice is limited, and the payoff from fine-grained mental-state inference declines (McNamara et al., 2019; Curtin et al., 2020).

Still, high kinship intensity does not eliminate strategic agency. Ethnography documents deliberate manipulation of kinship ties for personal or political advantage (Glick, 1967; Jakoubek & Budilová, 2006; Chagnon, 2017), showing that mental-state use remains situationally valuable. Among Turkana cattle raiders, for example, accusations of desertion are adjudicated through explicit discussion of motives, such as cowardice, before punishment is imposed (Mathew & Boyd, 2011).

Collectivism and Group Coordination

Collectivist settings, in which groups share corporate responsibility for members' acts, often emphasize the impact of behavior on others. Correspondingly, they feature greater condemnation of harmful outcomes (Feinberg et al., 2019). When intergroup conflict is

possible, prioritizing observable outcomes over disputable intentions can reduce contestation ("A caused B's death" rather than "Was it murder or manslaughter?") and thus preserve harmony (Curtin et al., 2020). In our terms, collectivist contexts tend to down-weight exculpatory intent and demand stronger evidence before excusing actions, especially in intergroup settings.

Empirically, Barrett et al. (2016) found that participants in two Global North communities (Los Angeles, rural Ukraine) attended more to mental states than did participants in eight small-scale Global South societies marked by dense kinship and collectivism. Yet, as Barrett & Saxe (2021) noted, context mattered everywhere: all societies showed within-scenario variation, cautioning against a simple "mind-using versus mind-blind" divide.

Material Insecurity and Threat

Resource scarcity, conflict, and ecological instability often strengthen collectivism, reduce relational mobility, and promote tighter behavioral norms (Jackson et al., 2020; Oishi & Komiya, 2017). Threat likewise increases moral harshness: scarcity heightens condemnation (Pitesa & Thau, 2014), and high-crime contexts amplify perceived responsibility (Tetlock et al., 2007). Severe harms lead observers to infer greater intentionality or negligence (Nobes & Martin, 2022), and sustained threat may have the same effect - tilting judgment toward outcomes and requiring stronger evidence before excusing.

Such tendencies align with the rationale of strict-liability laws, which focus on outcomes when the social cost of error is high. Barrett & Saxe (2021) note that, across societies, actions posing serious risks to others are often condemned independently of intent. Even within societies, vulnerability modulates this weighting: those most threatened - physically weaker individuals or those with high parental or mating investment - tend to condemn transgressions more strongly (DeScioli et al., 2014; Kurzban et al., 2010), sometimes inferring malicious intent where others do not (Liao et al., 2018; Tannenbaum et al., 2007).

Purity Violations

Taboos governing food and sex are often said to be judged with little or no regard to intent, perhaps because disgust-based mechanisms, that evolved to protect the individual from infection and suboptimal mate selection, treat violators as if they were contaminated – a threatening property that is independent of intent (Fessler & Navarrete, 2003; Young & Saxe, 2011). This possibility is particularly noteworthy given that numerous authors assert that the centrality of purity in moral values varies widely across cultures (Graham et al., 2011).

Early studies reported lower intent sensitivity for disgusting versus merely harmful acts (e.g., Chakroff et al., 2016; Young & Saxe, 2011; Young & Tsoi, 2013). More recent work finds that when intent information is matched across conditions, mental states play an equal role in judgements of disgusting and harmful acts (Kupfer et al., 2020; Kollareth & Russell, 2022). Even in Iran, a culture in which purity looms large (Atari et al., 2020), mental states play an equal role in judgements of disgusting and harmful acts (Sojoudi & Kupfer, in press). These more recent findings are understandable if one considers the logic of the claim that purity violations are insensitive to mental state information: that contaminating individuals pose a threat regardless of intent (Chakroff et al., 2013), or, similarly, that unintentional disgusting acts nevertheless elicit disgust (Russell & Giner-Sorolla, 2011). However, the question is not whether *the experience of disgust* is insensitive to intent, but whether *the moral judgement of the disgusting act* is sensitive to intent. In this way, disgusting moral violations may be no different from violations that elicit other emotions. Consider witnessing a person stabbed by another with a knife. We would experience horror regardless of information about the stabber's intent, but our moral judgement of the stabber would be strongly influenced by that information.[1]

Summary

Relational mobility, collectivism, and threat systematically shape when mental-state inference is deployed, how strongly it is weighted relative to outcomes, and which cues, intent, care, or foresight, receive emphasis. These influences operate in all societies, not just those with explicit opacity-of-mind norms. Socially transmitted beliefs and institutions that align with these facultative tendencies, such as favoring outcome-based deterrence under threat, are more likely to persist and spread. Cultural differences thus reflect facultative tuning of a shared moral architecture rather than fundamentally different kinds of minds. In the next section, we draw these strands together into an evolutionary framework that specifies how and why this facultative system adjusts the use of mental-state information across contexts.

Functional, Facultative, and Flexible: An Account of Mental-State Inference in Moral Judgment

[1] Some researchers argue that, unlike disgust (or horror), anger reactions to violations are sensitive to intent (Young and Tsoi, 2013). However, that is likely because anger is an output of the moral judgement of the act rather than a reaction to the act itself.

The cross-cultural patterns described above mirror those observed across judgment types, developmental stages, cognitive constraints, and legal systems. In each case, the weight given to mental states shifts predictably with the informational, social, and ecological conditions under which judgments are made. This variability is not noise but the product of an adaptive system that regulates the deployment and weighting of mental-state inference according to its expected benefits and costs.

Across the preceding sections, we have shown that sometimes people judge primarily on mental states, and at other times on outcomes. Under cognitive load, severe threat, or strict norms, judgments become more outcome-focused; when time, resources, and social stability permit, mental-state evidence is attended to more. We offer an integrative evolutionary account that explains these shifts, emphasizing that moral cognition serves multiple functions, and that both mind-contingent and effectively mind-blind processes can be adaptive under different conditions.

Partner Choice, Partner Control, and the Adaptive Value of Mindreading

One important function of moral psychology is partner choice: identifying trustworthy individuals for future collaboration. Because cooperation carries the danger of exploitation or abandonment, the ability to infer intentions helps people discern who is reliable (Barclay, 2016; Noë & Hammerstein, 1994). Accordingly, moral judgments often hinge on whether harm was intentional or accidental (Cushman, 2008), a distinction that preserves beneficial relationships and avoids unnecessary retaliation (Baumard et al., 2013; Landy & Uhlmann, 2018).

A second function is partner control through punishment (Barclay & Raihani, 2016). When alternative partners are unavailable, deterrence becomes paramount: imposing costs on wrongdoers discourages future harms (Raihani & Bshary, 2019). From this standpoint, detailed mental-state inference is not always necessary - punishing outcomes alone can suffice to signal that harm will not be tolerated (Martin & Cushman, 2015). If inferring intent is uncertain or costly, outcome-based punishment can be more efficient. Strict-liability rules exemplify this principle by simplifying blame assignment (Cappelletti, 2022).

The Role of Reputation Management

Moral judgments also serve a reputational function (Kurzban et al., 2007). Condemning or punishing wrongdoers can enhance one's own moral standing (Kupfer & Tybur, 2023), especially when the transgressor is widely disliked. Swift, uncompromising condemnation can signal loyalty and moral conviction even if it

overlooks exculpatory intent. Conversely, in contexts where fairness and impartiality are prized, careful evaluation of intentions can signal reliability and trustworthiness (Malle et al., 2022; Baumard et al., 2013). Reputation management therefore alters the cost-benefit calculus of how extensively and carefully mental states are inferred. In combination, these partner-choice, partner-control, and reputational functions explain why moral cognition is both facultative and flexible.

Facultative and Flexible Mechanisms of Moral Judgment

Together, these functions entail that moral judgment relies on facultative and flexible mechanisms. Rather than a single monolithic process, moral cognition operates through multiple, functionally specialized subsystems. We refer to this organization as the Facultative Mentalizing System (FMS) - a functional architecture, not a discrete neural module.

Facultative deployment governs whether and how extensively mental-state inference is engaged. It includes not just the decision to consider an agent's intentions or beliefs, but also the depth of that consideration - the amount of mental-state evidence observers seek before forming a judgment. The FMS integrates several streams of input: outcome severity; the relationship value of wrongdoer and victim (kin, ally, stranger, outgroup); threat level; normative standards for punishment; and reputational incentives tied to audience expectations. A further input, behavioral and contextual evidence such as whether precautions were taken or attempts at prevention made, is incorporated only to the extent that mentalizing is engaged. From this evidence, people infer mental states such as intention, foresight, or care.

The system performs two roles. First, it integrates contextual streams to estimate the expected value of recruiting mental-state inference. Second, if engaged, it evaluates and flexibly weights the resulting inferences. This flexibility has two forms: (i) weighting mental-state inferences relative to outcomes, and (ii) selectively amplifying exculpatory or inculpatory cues depending on social context, for instance, excusing an ally or condemning a rival.

Together, these processes yield moral judgments that vary not only in whether mental states matter but also in how they matter, judgments located along a continuum from mind-contingent to effectively mind-blind.

Cultural and Ecological Influences

The same facultative mechanisms help explain cultural and legal variation. Under resource scarcity or collective danger, many societies adopt norms that favor outcome-based judgments to deter harm regardless of intent. Legal systems often parallel these tendencies,

using strict-liability rules to simplify deterrence of risky behaviors. By contrast, where due process and fairness are highly prized, institutions emphasize mens rea requirements and encourage careful assessment of intentions and beliefs. This variability is possible because the underlying psychology which is able to facultatively deploy and flexibly weight mental-state information is itself adaptive to ecological and cultural demands. Societies also differ in how they evaluate judges: in audiences that value uncompromising condemnation, ignoring mental states can raise prestige; in audiences that prize impartiality and empathy, intent-sensitive judgments confer reputational benefits.

Conclusion: Functional, Facultative, and Flexible

Moral judgment serves several functions: deterring harmful acts through punishment, identifying trustworthy partners, and managing reputation. These aims sometimes converge but often compete, producing systematic variation in how mental states are used. The same adaptive system underlies these shifts: a facultative mechanism that deploys mental-state inference when the expected benefits of accuracy and coordination outweigh its cognitive and social costs, and weights that evidence flexibly against outcomes and contextual cues. From childhood to old age, from dyadic morality to legal institutions, and across cultural ecologies, people adjust whether and how they use mental-state information in ways that balance the costs of inference against the risks of error. The continuum from mind-contingent to mind-blind judgment is not a flaw of moral cognition but an evolved solution to the recurring problem of deciding when minds matter.

Acknowledgments

We thank Jared Piazza and Justin Landy for helpful feedback. This research was supported by the U.S. Air Force Office of Scientific Research under Award #FA9550-23-1-0694 to DF and TK. The funder's website is https://www.afrl.af.mil/AFOSR/. The funder had no role in the design, contents, decision to publish, or preparation of the manuscript.

References

Abrams, D., Randsley de Moura, G., & Travaglino, G. A. (2013). A double standard when group members behave badly: Transgression credit to ingroup leaders. *Journal of Personality and Social Psychology*, *105*(5), 799. https://doi.org/10.1037/a0033600

Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fmri investigation. *Journal of Experimental Social Psychology*, *46*(6), 1142–1148. https://doi.org/10.1162/jocn.2009.21187

Alicke, M. D., & Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, *39*(9), 2089–2105. https://doi.org/10.1111/j.1559–1816.2009.00517.x

American Law Institute. (1962). *Model penal code*. ALI.

American Law Institute. (1981). *Restatement (second) of the law of contracts*. ALI.

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755–1762. https://doi.org/10.1177/0956797613480507

Anderson, R. A., Crockett, M. J., & Pizarro, D. A. (2020). A theory of moral praise. *Trends in Cognitive Sciences*, *24*(9), 694–703. https://doi.org/10.1016/j.tics.2020.06.008

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of "theory of mind"*. Psychology Press.

Atari, M., Graham, J., & Dehghani, M. (2020). Foundations of morality in iran. *Evolution and Human Behavior*, *41*(5), 367–384. https://doi.org/10.1016/j.evolhumbehav.2020.07.014

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37–49. https://doi.org/10.1002/cd.96

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology*, *7*, 33–38. https://doi.org/10.1016/j.copsyc.2015.10.012

Barclay, P., & Raihani, N. (2016). Partner choice versus punishment in human cooperation and sociality. *Evolution and Human Behavior*, *37*(6), 345–357.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Barrett, H. C., & Saxe, R. R. (2021). *Are some cultures more mind-minded in their moral judgements than others? philosophical transactions of the royal society b, 376(1838), 20200288*.

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., … Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, *113*(17), 4688–4693. https://doi.org/10.1073/pnas.1522070113

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78.

Binder, G. (2008). The culpability of felony murder. *Notre Dame Law Review*, *83*, 965–1060.

Bocian, K., Cichocka, A., & Wojciszke, B. (2021). Moral tribalism: Moral judgments of actions supporting ingroup interests depend on collective narcissism. *Journal of Experimental Social Psychology*, *93*, 104098.

Buon, M., Dupoux, E., Jacob, P., Chaste, P., Leboyer, M., & Zalla, T. (2013). The role of causal and intentional judgments in moral reasoning in individuals with high functioning autism. *Journal of Autism and Developmental Disorders*, *43*, 458–470.

Buon, M., Seara-Cardoso, A., & Viding, E. (2016). *Why (and how) should we study the interplay between emotional arousal, theory of mind, and inhibitory control to understand moral cognition? psychonomic bulletin & review, 23, 1660–1680*.

Cappelletti, M. (2022). *Justifying strict liability: A comparative analysis in legal reasoning*. Oxford University Press.

Chagnon, N. A. (2017). Manipulating kinship rules: A form of male yanomamö reproductive competition. In Adaptation and human behavior (pp. 115–132). Routledge.

Chakroff, A., Dungan, J., & Young, L. (2013). *Harming ourselves and defiling others: What determines a moral domain? PLOS ONE, 8(9), e74434*.

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476–484.

Child, J., & Ormerod, D. (2023). *Smith, hogan and ormerod's essentials of criminal law*. Oxford University Press.

Convention on Damage Caused by Foreign Aircraft to Third Parties on the Surface. (1952).

Convention on International Liability for Damage Caused by Space Objects. (1972).

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235.

Cosmides, L., & Tooby, J. (2006). Evolutionary psychology, moral heuristics, and the law. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the Law* (pp. 181–212). MIT Press.

Curtin, C. M., Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., … Henrich, J. (2020). Kinship intensity and the use of mental states in moral judgment across societies. *Evolution and Human Behavior*, *41*(5), 415–429.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292.

Cushman, F. (2014). Punishment in humans: From intuitions to institutions. *Philosophy Compass*, *10*(2), 117–133.

Cutter v. Powell, 101 Eng. Rep. 573 (1795) (Eng.).

Dempsey, E. E., Moore, C., Johnson, S. A., Stewart, S. H., & Smith, I. M. (2020). Morality in autism spectrum disorder: A systematic review. *Development and Psychopathology*, *32*(3), 1069–1085.

Denning, A. T. (1953). *The changing law*. Steven and Sons.

DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? moral judgments follow the money. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20142112.

Donoghue v Stephenson, [1932] AC 562.

Dubois, J. (1993). Meaning without intention: Lessons from divination. In J. Hill & J. Irvine (Eds.), *Responsibility and Evidence in Oral Discourse* (pp. 48–71). Cambridge University Press.

Ekerim-Akbulut, M., Selçuk, B., Slaughter, V., Hunter, J. A., & Ruffman, T. (2020). In two minds: Similarity, threat, and prejudice contribute to worse mindreading of outgroups compared with an ingroup. *Journal of Cross-Cultural Psychology*, *51*(1), 25–48.

Epstein, L. A. (2006). There are not bad dogs, only bad owners: Replacing strict liability with a negligence standard in dog bite cases. *Animal L.*, *13*, 129.

Feinberg, M., Fang, R., Liu, S., & Peng, K. (2019). A world of blame to go around: Cross-cultural determinants of responsibility and punishment judgments. *Personality and Social Psychology Bulletin*, *45*(4), 634–651.

Fessler, D. M. T., & Navarrete, C. D. (2003). Meat is good to taboo: Dietary proscriptions as a product of the interaction of psychological mechanisms and social processes. *Journal of Cognition and Culture*, *3*(1), 1–40.

Fittipaldi, S., Armony, J. L., García, A. M., Migeot, J., Cadaveira, M., Ibáñez, A., & Baez, S. (2023). Emotional descriptions increase accidental harm punishment and its cortico-limbic signatures during moral judgment in autism. *Scientific Reports*, *13*(1), 1745.

Flume, J. W. (2022). Strict liability in austrian and german law: On the concept of strict liability in the age of technological advancement. *Journal of European Tort Law*, *12*(3), 205–220.

Forbes, R. C., & Stellar, J. E. (2022). When the ones we love misbehave: Exploring moral processes within intimate bonds. *Journal of Personality and Social Psychology*, *122*(1), 16.

Frith, U. (2004). Emmanuel miller lecture: Confusions and controversies about asperger syndrome. *Journal of Child Psychology and Psychiatry*, *45*(4), 672–686.

Frith, U., & Happé, F. (1994). *Autism: Beyond "theory of mind." cognition, 50(1–3), 115–132*.

Garcia-Molina, I., Clemente-Estevan, R. A., & Andrés-Roqueta, C. (2021). Investigating moral judgements in autistic children: Integrating the observer's and the speaker's mind. *Journal of Cognitive Psychology*, *33*(1), 87–101.

Gaunt, R., Leyens, J. P., & Demoulin, S. (2002). Intergroup relations and the attribution of emotions: Control over memory for secondary emotions associated with the ingroup and outgroup. *Journal of Experimental Social Psychology*, *38*(5), 508–514.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1998). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733–740.

Glick, L. B. (1967). The role of choice in gimi kinship. *Southwestern Journal of Anthropology*, *23*(4), 371–382.

Gneezy, A., & Fessler, D. M. (2012). Conflict, sticks and carrots: War increases prosocial punishments and rewards. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1727), 219–223.

Goudkamp, J., & Nolan, D. (2020). *Winfield and jolowicz on tort (20th ed.)*. Sweet & Maxwell.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366.

Gray, K., DiMaggio, N., Schein, C., & Kachanoff, F. (2023). The problem of purity in moral psychology. *Personality and Social Psychology Review*, *27*(3), 272–308.

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*(3), 505–520.

Haidt, J. (2012). Moral psychology and the law: How intuitions drive reasoning, judgment, and the search for evidence. *Alabama Law Review*, *64*, 867–880.

Hart, H. M. (1958). The aims of the criminal law. *Law and Contemporary Problems*, *23*(3), 401–441.

Holmes, O. (1881). *The common law*. Little, Brown & Co.

Jackson, J. C., Gelfand, M., & Ember, C. R. (2020). A global analysis of cultural tightness in non-industrial societies. *Proceedings of the Royal Society B*, *287*(1930), 20201036.

Jakoubek, M., & Budilová, L. (2006). Kinship, social organisation and genealogical manipulations in gypsy osadas in eastern slovakia. *Romani Studies*, *16*(1), 63–82.

Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., & Charman, T. (2010). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, *3*(2), 98–109.

Jones, O. D. (2001). Proprioception, non-law, and biolegal history. *Florida Law Review*, *53*, 831–874.

Kagan, J. (1981). *The emergence of morality in young children*. University of Chicago Press.

Keating, G. C. (2014). Strict liability wrongs. In J. Oberdiek (Ed.), *Philosophical Foundations of the Law of Torts* (pp. 292–311). Oxford University Press.

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally relevant theory of mind. *Cognition*, *119*(2), 197–215.

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, *182*, 331–348.

Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190–194.

Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*(2), 205–224.

Kollareth, D., & Russell, J. A. (2022). When judging purity norm violations, the perpetrator's intention matters. *European Journal of Social Psychology*, *52*(5–6), 931–943.

Kupfer, T. R., & Tybur, J. M. (2023). Third-party punishers who express emotions are trusted more. *Proceedings of the Royal Society B*, *290*(2005), 20230916.

Kupfer, T. R., Inbar, Y., & Tybur, J. M. (2020). Reexamining the role of intent in moral judgements of purity violations. *Journal of Experimental Social Psychology*, *91*, 104043.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*(2), 75–84.

Larkin Jr, P. J. (2014). Strict liability offenses, incarceration, and the cruel and unusual punishments clause. *Harv. J.L. & Pub. Pol'y*, *37*, 1065.

Lee, J., & Holyoak, K. J. (2020). "but he's my brother": The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, *48*, 158–170.

Liao, Y., Sun, Y., Li, H., Deák, G. O., & Feng, W. (2018). Intensity of caring about an action's side-effect mediates attributions of actor's intentions. *Frontiers in Psychology*, *9*, 1329.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, *5*(2), 147470490700500203.

Lieberman, D., & Patrick, C. (2018). *Objection: Disgust, morality, and the law*. Oxford University Press.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186.

Malle, B. F., Guglielmo, S., Voiklis, J., & Monroe, A. E. (2022). Cognitive blame is socially shaped. *Current Directions in Psychological Science*, *31*(2), 169–176.

Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development*, *41*, 46–64.

Margoni, F., Geipel, J., Hadjichristidis, C., & Surian, L. (2019). The influence of agents' negligence in shaping younger and older adults' moral judgment. *Cognitive Development*, *49*, 116–126.

Margoni, F., Geipel, J., Hadjichristidis, C., & Surian, L. (2021). Inequity aversion in old age: An outcome bias in older adults' socio-economic decisions. *Cognitive Development*, *58*, 101016.

Margoni, F., Geipel, J., Hadjichristidis, C., Bakiaj, R., & Surian, L. (2023). Age-related differences in moral judgment: The role of probability judgments. *Cognitive Science*, *47*(9), e13345.

Margoni, F., Guglielmetti, G., & Surian, L. (2019). Brief report: Young children with autism can generate intent-based moral judgments. *Journal of Autism and Developmental Disorders*, *49*(12), 5078–5085.

Martin, J. W., & Cushman, F. (2015). To punish or to leave: Distinct cognitive processes underlie partner control and partner choice behaviors. *PLOS ONE*, *10*(4), e0125193.

Martin, J. W., & Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, *147*, 133–143.

Martin, J. W., Buon, M., & Cushman, F. (2021). The effect of cognitive load on intent-based moral judgment. *Cognitive Science*, *45*(4), e12965.

Martin, J. W., Buon, M., & Cushman, F. (2021). The effect of cognitive load on intent-based moral judgment. *Cognitive Science*, *45*(4), e12965.

Mathew, S., & Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, *108*(28), 11375–11380.

McKee, P., Kim, H. E., Tang, H., Everett, J. A., Chituc, V., Gibea, T., … & Sinnott-Armstrong, W. (2024). Does it matter who harmed whom? a cross-cultural study of moral judgments about harm by and to insiders and outsiders. *Current Psychology*, *43*(9), 7997–8007.

McLoughlin, N., & Over, H. (2017). Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychological Science*, *28*(10), 1503–1509.

McManus, R. M., Kleiman-Weiner, M., & Young, L. (2020). What we owe to family: The impact of special obligations on moral judgment. *Psychological Science*, *31*(3), 227–242.

McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. *Intent Across Societies: How Cultural Models Of Mind Shape Moral Reasoning. Cognition*, *182*, 95–108.

Mikhail, J. (2009). Is the prohibition of homicide universal? evidence from comparative criminal law. *Brooklyn Law Review*, *75*, 497–515.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge University Press.

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology*, *116*(2), 215.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557.

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, *108*(7), 2688–2692.

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell L. Rev.*, *97*, 255.

Nettleship v Weston, [1971] QB 691.

Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science*, *25*(3), 648–655.

Niemi, L., & Young, L. (2016). When and why we see victims as responsible: The impact of ideology on attitudes toward victims. *Personality and Social Psychology Bulletin, 42*(9), 1227–1242.

Nobes, G., & Martin, J. W. (2022). They should have known better: The roles of negligence and outcome in moral judgements of accidental actions. *British Journal of Psychology*, *113*(2), 370–395.

Nobes, G., Panagiotaki, G., & Engelhardt, P. E. (2017). The development of intention-based morality: The influence of intention salience and recency, negligence, and outcome on children's and adults' judgments. *Developmental Psychology*, *53*(10), 1895.

Nobes, G., Panagiotaki, G., & Martin, J. W. (2023). Moral luck and the roles of outcome and negligence in moral judgments. *Journal of Experimental Social Psychology*, *106*, 104456.

Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology*, *104*(4), 382–397.

Noë, R., & Hammerstein, P. (1994). Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology*, *35*, 1–11.

Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, *49*(2), 100–110.

Oishi, S., & Komiya, A. (2017). Natural disaster risk and collectivism. *Journal of Cross-Cultural Psychology*, *48*(8), 1263–1270.

Owen, D. G. (1997). Philosophical foundations of fault in tort law. In D. G. Owen (Ed.), *The Philosophical Foundations of Tort Law* (pp. 201–228). Oxford University Press.

Park, B., Vepachedu, S., Keshava, P., & Minns, S. (2022). Culture, theory-of-mind, and morality: How independent and interdependent minds make moral judgments. *Biological Psychology*, *174*, 108423.

Patil, I., & Trémolière, B. (2021). Reasoning supports forgiving accidental harms. *Scientific Reports*, *11*(1), 14418.

Patrick, C. (2023). Evolution is the source, and the undoing, of natural law. *Evolution and Human Behavior*, *44*(3), 175–183.

Patrick, C. J., & Lieberman, D. (2017). Not from a wicked heart: Testing the assumptions of the provocation doctrine. *Nevada Law Journal*, *18*, 33–59.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163–177.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology*, *46*(2), 530–544.

Piaget, J. (1932). *The moral judgment of the child*. Harcourt, Brace.

Pitesa, M., & Thau, S. (2014). A lack of material resources causes harsher moral judgments. *Psychological Science*, *25*(3), 702–710.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science*, *14*(3), 267–272.

R v Lipman, [1970] 1 QB 152.

Raihani, N. J., & Bshary, R. (2019). Punishment: One tool, many uses. *Evolutionary Human Sciences*, *1*, e12.

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, *1*(4), 223–235.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881.

Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, *86*(4), 530–544.

Restatement of Torts, Second (U.S.).

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, *81*(2), 407–420.

Robinson, P. H. (2021). Criminal law's core principles. *Washington University Jurisprudence Review*, *14*(1), 153–218.

Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. Westview.

Rossano, F., Rakoczy, H., & Tomasello, M. (2011). Young children's understanding of violations of property rights. *Cognition*, *121*(2), 219–227.

Rumsey, A. (2013). Intersubjectivity, deception and the "opacity of other minds": Perspectives from highland new guinea and beyond. *Language & Communication*, *33*(3), 326–343.

Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion*, *11*(2), 233.

Russell, P. S., & Piazza, J. (2015). Consenting to counter-normative sexual acts: Differential effects of consent on anger and disgust as a function of transgressor or consenter. *Cognition and Emotion, 29*(4), 634–653.

Rylands v Fletcher, (1868) LR 3 HL 330.

Sarin, A., & Cushman, F. (2024). One thought too few: An adaptive rationale for punishing negligence. *Psychological Review*, *131*(3), 812.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The church, intensive kinship, and global psychological variation. *Science*, *366*(6466), eaau5141.

Schwartz, F., Djeriouat, H., & Trémolière, B. (2022). Agents' moral character shapes people's moral evaluations of accidental harm transgressions. *Journal of Experimental Social Psychology*, *102*, 104378.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249.

Searle, J. R. (2001). *Rationality in action*. MIT Press.

Shen, F. X., Hoffman, M. B., Jones, O. D., & Greene, J. D. (2011). Sorting guilty minds. *New York University Law Review*, *86*, 1306–1360.

Sojoudi, A., & Kupfer, T. R. (under review). Intent matters for moral judgements of purity violations even in Iran. Manuscript under review.

Szekely, A., Lipari, F., Antonioni, A., Paolucci, M., Sánchez, A., Tummolini, L., & Andrighetto, G. (2021). Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, *12*(1), 5452.

Sznycer, D., & Patrick, C. (2020). The origins of criminal law. *Nature Human Behaviour*, *4*, 506–516.

Tannenbaum, D., Ditto, P. H., & Pizarro, D. A. (2007). *Different moral values produce different judgments of intentional action*. Unpublished manuscript, University of California–Irvine.

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*(6), 1249–1254.

Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., … & Rescober, P. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, *43*(2), 195–209.

Tettenborn, P. A. (Ed.). (2023). *Clerk & lindsell on torts (24th ed.)*. Sweet & Maxwell.

Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., … Ferreira, C. M. (2018). Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proceedings of the National Academy of Sciences*, *115*(29), 7521–7526.

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, *64*, 231–255.

Turpin, M. H., Walker, A. C., Fugelsang, J. A., Sorokowski, P., Grossmann, I., & Białek, M. (2021). The search for predictable moral partners: Predictability and moral (character) preferences. *Journal of Experimental Social Psychology*, *97*, 104196.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81.

Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*(2), 326–334.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124–130.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*(8), 689–690.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.

Waytz, A., Young, L. L., & Ginges, J. (2014). Motive attribution asymmetry for love vs. *Hate Drives Intractable Conflict. Proceedings Of The National Academy Of Sciences*, *111*(44), 15687–15692.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

White, G. E. (2003). *Tort law in america: An intellectual history*. Oxford University Press.

Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, *168*, 154–163.

Wyman, E., Rakoczy, H., & Tomasello, M. (2009). Normativity and context in young children's pretend play. *Cognitive Development*, *24*(2), 146–155.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*(4), 1912–1920.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214.

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, *7*(8), 585–604.

Zalla, T., Clement, A., & Leboyer, M. (2011). *Is theory of mind delayed in non-retarded children with autism? cognitive development, 26(2), 138–155*.

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, *67*(5), 2478–2492.