Human beings live complex social lives, composed of various types of relationships across nested social hierarchies, all structured by rights, rules, and obligations. However, selfish goals persist, and keeping individuals' goals in line with community interests has become the primary challenge of modern morality. To meet this challenge, human societies have developed two major social-cultural tools: a vast network of rules, norms, and values (Sripada & Stich, 2006; Ullmann-Margalit, 1977) and complex social practices of norm enforcement, such as blaming, praise, apology, and reconciliation (Semin & Manstead, 1983).

This kind of social-cultural morality has to be taught, learned, and enforced by community members, even by the youngest among them (Göckeritz, Schmidt, & Tomasello, 2014). Acquiring norms likely benefits from early-appearing preferences for prosocial agents over antisocial agents (Hamlin, 2014), but socially mature moral capacities rely heavily on nonmoral capacities: those of social cognition (Guglielmo, Monroe, & Malle, 2009).

Social cognition encompasses a hierarchy of interdependent concepts, processes, and skills that allow individuals to perceive, understand, and—most important for the present topic—evaluate one another. For example, norm enforcers infer the mental processes that generated a transgressive behavior (e.g., motive, belief, intention) before blaming the transgressor (Malle, Guglielmo, & Monroe, 2014). The transgressor must likewise infer the mental processes that generated the social act of blaming (e.g., the norm enforcer's goals, knowledge, and power of enforcing sanctions) when deciding to deny or admit, maintain or correct the norm-violating behavior.

These, then, are the phenomena this chapter aims to illuminate. We show how the elements of social cognition ground people's moral cognition and how social and moral cognition together guide the social regulation of behavior by moral norms. We aim to identify the concepts, mechanisms, and practices that go into making various kinds of moral judgments and the forms and functions of socially expressing those judgments.

Historical Context

The most prominent debates in moral philosophy grapple with dichotomies. Perhaps the oldest of these concerns the relative influences of reason and passion on human behavior (Hume, 1751/1998; Kant, 1785/2012). Moral psychology, too, has been heavily influenced by this dichotomy. During an early phase, scholars expressed great confidence in the human capacity to reason about moral matters—albeit a capacity that needs time to develop (Piaget, 1932; Kohlberg, 1981). During a later phase, scholars expressed sometimes fierce skepticism toward such reasoning capacities and offered emphatic claims about the primacy of affect in moral judgment (Alicke, 2000; Greene, 2008), about people's inability to access the cognitive basis of their judgments (Nisbett & Wilson, 1977; Haidt, 2001), and about the many biases from which these judgments suffer (Ditto, Pizarro, & Tannenbaum, 2009).

Dichotomies often suffer from exaggerations and simplifications. We hope to present a framework that goes beyond extreme positions and relies instead on theoretical analysis, existing empirical evidence, and predictions about new phenomena. We believe that drawing a line and designating two opposing sides—reason versus passion, cognition versus emotion, deliberation versus intuition—is an unproductive way to tackle a multifaceted phenomenon. We should, rather, survey the landscape and acknowledge the complex terrain of social life so as to discover the different psychological adaptations and social practices that have allowed people to navigate the terrain—imperfectly, but not in as bumbling and blundering a way as is sometimes portrayed. What enables such adaptive navigation, we try to show, is the interactive system of moral cognition, social cognition, and social regulation.

This system is schematically illustrated in Figure 12.1, in which the parts both inform one another (e.g., mental state inferences informing a blame judgment) and also justify one another (e.g., a wrongness judgment providing justification for an act of social regulation).

Two aspects of this schematic deserve comment. First, when we use the term social cognition, we are not pitching our tent on the "reason" side of a dichotomy but rather conceiving of social cognition as a large toolbox that contains both fast and automatic mechanisms as well as slow and controlled mechanisms, both automatic and
The moral domain is that of regulating individual behavior in the context of community interests. Rules, norms, and values set the standards that, if fulfilled, serve community goals and allow the community and its individuals to succeed. The broader literature sometimes distinguishes moral from conventional rules or moral from social norms (Brennan, Eriksson, Goodin, & Southwood, 2013; Kohlberg, 1981). But for many purposes, it is best to assume a continuum of norms—defined as standards and instructions that guide people in what they should do. We can then identify prototypes at each end. Moral norms, on one end, are part of a hierarchy in which moral “principles” and “values” are the most abstract instructions; social-conventional norms, at the other end, can often stand alone in regulating just one particular behavior or in solving a coordination problem. What the elements of this continuum have in common is that, in representing an instruction as a norm (as opposed to a goal or habit), people keenly take into account that (1) a sufficient number of individuals in the community in fact follow the instruction, and (2) a sufficient number of individuals in the community expect and demand of each other to follow the instruction and may be willing to enforce it through sanctions (Bicchieri, 2006; Brennan et al., 2013).

We can now conceptualize moral cognition as the set of capacities that allow people to properly engage with social and moral norms. People have to (1) learn, store, activate, and deploy norms; (2) make judgments (e.g., of permisibility, wrongness, blame) about these norms; (3) make decisions in light of these norms; and (4) communicate about the norms and their violations (e.g., prescribe, justify, apologize).1

How Is Social Cognition Involved in Moral Cognition?

What is social cognition? We endorse an inclusive definition that subsumes under the term all conceptual and cognitive tools that serve the overarching goal of making sense of other human agents. Figure 12.2 displays many of these tools arranged in tree-like hierarchy (for a glossary and detailed discussion, see Malle, 2008, 2015). On the bottom are those that have evolved earlier in phylogeny, develop earlier in ontogeny, and are generally simpler and faster processes; on the top are those that have evolved more recently, develop later in childhood, and are generally more complex and slower processes. The tools often rely on the output of tools below them, and in concert these tools perform important tasks in social life, such as explanation, prediction, and moral judgment (depicted outside the tree itself). Moreover, several of the processes at once presuppose and shape fundamental concepts, such as intentionality, belief, desire, and emotion categories.

Against this background it is now easy to illustrate how social cognition supports and interacts with the four capacities of moral cognition.

In norm learning, social cognition contributes some of the learning mechanisms: Mimicry and imitation provide powerful tools of adopting norms through action, and face processing and goal identification allow people to read others’ evaluations of a given behavior and thereby infer the norms that the behavior conformed to or violated. For example, a scowl toward somebody who asks a new acquaintance too many private questions can teach and enforce norms of privacy and autonomy.

Moral communication, finally, includes such phenomena as expressing moral judgments either to the alleged violator or to another community member (Dersley & Wootton, 2000; Traverso, 2009); negotiating blame through justification and excuses (Antaki, 1994); and apology, compensation, or forgiveness to repair social estrangement after a norm violation (McKenna, 2012; Walker, 2006). People rely on mental state inferences during communicative interactions, and especially during social–moral interactions, to accurately assess the other’s goals and knowledge, because the stakes of maintaining relationships are high and under the threat of sanctions. Trait inferences may be formed through observation or gossip, especially when norm violators do not respond to social regulation attempts by their community. Also, low-level tools of gaze and face processing, empathy, and goal inference are needed to gauge the honesty of justifications, the genuineness of apologies, and the seriousness of threatened sanctions.

How Does Social Regulation Interact with Moral Cognition?

We claimed earlier that heeding the social regulatory function of moral cognition can benefit our understanding of how moral cognition itself operates. We now illustrate...
The accuracy or bias of a given moral judgment is difficult to measure, because the laboratory rarely offers an objective criterion for the correct judgment. Typically, researchers offer information to participants that they “should not” take into account, and when some of them do, a “bias” is diagnosed. For example, many researchers have argued that outcome severity, the motives and character of the norm violator, or the likeability of the victim must not be part of an unbiased moral judgment. But it is unclear who gets to decide, and on what basis, what people should or should not take into account (Malle et al., 2014; Nadler, 2012). Moreover, the potential arbiters, “philosophers, legal theorists and psychologists” (Alicke, 2008, p. 179), often do not agree with one another.

In the absence of objective criteria, an appealing alternative is to consider a moral judgment’s function of regulating social behavior as a suitable standard—the socially shared criteria that people use to accept, question, criticize, or reject moral judgments. For example, what do people accept as the grounds of intense blame? They consider that the behavior violated an important norm, that the violation was intentional, that the agent had no justifying reasons to perform the behavior, and so forth (Malle et al., 2014). When people reject intense blame they do so when the behavior violated a merely insignificant norm, when the violation was unintentional and unavoidable but the norm enforcer treated it as if it were intentional, and so forth. Bias is then diagnosed when norm enforcers overblame or underblame relative to what is acceptable in the community (Kim, Voiklis, Cusimano, & Malle, 2015).

These standards of blame put pressure on people to keep their biases in check. Severe violations sometimes elicit powerful emotional responses that can lead to premature accusations or unfair punishment; further, an observer’s quick moral evaluation sometimes taints subsequent inferences about whether the violation was intentional, justifiably, or preventable (Alicke, 2000; Knobe, 2010). Nevertheless, community members help correct these expressions of premature, biased, or inaccurate moral judgments by challenging those who blurt out allegations and by demanding warrant from those who overblame, thereby calming and slowing the processes of accusation and punishment. Communities could not survive if their members blamed and punished one another without evidence or without differentiating between, say, mild and severe, intentional and unintentional violations. The regulatory functions of moral judgment, and the required warrant for such judgments, therefore push these judgments to be more reasonable, accurate, and fair, by the standards of the community in which they occur.

Relating Social Cognition, Moral Cognition, and Social Regulation

We can now offer a more detailed schematic of the relationships between social cognition, moral cognition, and social regulation. In choosing the pictorial language of a flow diagram (Figure 12.3), we naturally leave out some complexity, but it forces us to make explicit certain theoretical commitments, which can be tested experimentally.

The flow of processes begins with a negative event, which prompts the perceiver to assess whether the event was caused by an agent who violated norms. If “yes,” social cognitive processes analyze the violator’s mental states (including intentions and goals). This information feeds into moral judgment, which generates judgments about wrongness or blame. The outputs of moral and social cognition, along with preceding information about the event and the norms that were violated, feed into a decision about whether public moral criticism is warranted. If warrant exceeds threshold, the perceiver is likely to deliver public moral criticism (though many other considerations may inhibit criticism, such as role constraints, fear of retaliation, etc.). This moral criticism may prompt a timely change in the violator’s behavior or, if not, the perceivers may consider renewed criticism or alternative responses, including gossip or retreat.

The full step with which we break off the flow diagram conceals a more complex, finely tuned social dynamic between norm enforcers and norm violators: They negotiate levels of blame, meet accusation with justification, criticism with remorse, remorse with forgiveness, all in the service of rebuilding and maintaining social relationships (Walker, 2006).

Evidence

Empirical evidence for the social cognitive basis of moral judgment has been accumulating over the past several years. In many studies, lay people clearly rely on social cognitive inferences of intentionality when judging everyday moral actions (Lagnado & Channon, 2008) and when mastering fine distinctions between willingly, knowingly, intentionally, and purposefully violating a norm (Guglielmo & Malle, 2010)—distinctions that also inform legal classifications of negligence and recklessness. Likewise, lay people judge goal-directed harm as less permissible and more often as wrong than they judge harm as a side effect (Cushman & Young, 2011). Thus moral and legal distinctions overlap with (and perhaps derive from) more general-purpose social cognitive judgments. This derivative relationship is corroborated by results from functional magnetic resonance imaging and lesion studies showing that the processing involved in either social or moral judgment activate many of the same regions in the prefrontal cortex (Forbes & Grafman, 2010).

People's cognitive system also makes distinctions between types of moral judgments that vary by the objects they judge: Badness judges mere events, wrongness judges intentional actions, and blame judges an agent's specific relationship to a norm violation, whether intentional or unintentional (Malle et al., 2014; Monin, Pizarro, & Beer, 2007; Sher, 2006). These judgments also differ in their sensitivity to causal and mental state in-
formation (Cushman, 2008; Malle, Scheutz, Arnold, Voiklis, & Cushman, 2015), but experiments on the detailed causal processes that flow between social cognition and these differing judgments remain lacking. Experiments on social expressions of blame are also scarce. Nevertheless, initial work in our lab has demonstrated that, as with private judgments, people have a finely tuned map of public acts of moral criticism (Voiklis, Cushman, & Malle, 2014). For example, the rich vocabulary used by English speakers to describe such acts—ranging from chiding violators to lashing out at them—do not merely represent linguistic variations but pick out systematic features of the underlying moral judgment and of the social context. When participants assessed 28 acts (described by the most common verbs of moral criticism) on numerous properties of judgment and context, the first two dimensions of a principal components analysis were intensity of expression and direction of expression (toward the offender or toward others). Figure 12.4 depicts the quadrants of this space and four verbs that mark the prototypical acts in each quadrant. In a subsequent series of studies, we tested the hypothesis that people likely follow “norms of blaming” when scaling the intensity of moral criticism to the severity of transgressions (Kim et al., 2015). Indeed, when judging the appropriateness of various levels of moral criticism in response to a range of mild to severe transgressions, participants displayed a norm against “overblaming” (i.e., overly intense criticism for mild violations) but were more tolerant of “underblaming.”

**Individual and Situational Variability in Social and Moral Cognition**

So far, we have addressed social and moral cognition at the level of cognitive system components that exist in all neurotypical adults. Nevertheless, social cognitive performance can vary as a function of maturation, neurological damage, and psychopathology (Frith & Frith, 2003) and can also be due to motivation (Klein & Hodges, 2001) and task difficulty (Birch & Bloom, 2007). Often these deficits are presented as evidence that people are reflexively egocentric in their perception of other minds (Lin, Keysar, & Epley, 2010). An alternative interpretation is that people are dispositionally or situationally unprepared for attending to the full range of social information. In fact, preliminary evidence suggests that “warning up” social cognition with a practice task facilitates judgment of moral blame. So even though shallow processing and bias may predominate in states of disengagement, the correct situational cues can bring most individuals to their full social cognitive potential. Among these situational cues, the community’s demand for warrant in moral criticism (especially blame) must rank very high, but direct tests of this hypothesis remain lacking.

There is, however, evidence for the malleability and the social shaping of moral reasoning more generally. As with other forms of public reasoning (Crowell & Kuhn, 2012; Kuhn, Zillmer, Crowell, & Zavala, 2013), moral judgment can improve with practice and feedback. Much as habitual reliance on heuristics (confirmation seeking) can be overcome with deliberate practice (Kuhn, 2011), people might likewise overcome any habitual neglect of social cognitive information. Howe (1991), for example, showed in an experimental context that circuit judges adjusted their blame judgments to mitigating information twice as strongly as students did. Applying one’s social cognitive abilities might also be a matter of mindset. When induced to believe in the malleability, as opposed to the fixedness, of empathy, people appear more willing to expend empathic effort toward challenging targets (Schumann, Zaki, & Dweck, 2014). Moreover, people with a malleable mindset appear to seek out these challenges in order to improve their empathy; the challenge provides the learning opportunity, and the motivation to learn helps them meet that challenge.

Beyond skill learning, the vast developmental literature on changes in moral judgment and decision making support the claim of malleability. Gradual differentiation in moral cognition, according to our framework, is in good part the result of gradual differentiation in social cognition (Baird & Astington, 2004). For example, norm learning becomes more sophisticated as moral state inferences improve, and blame judgments become more sophisticated as the conceptual framework of mind grows. Specifically, as mental state concepts of belief and desire mature by ages 4–5 (Wellman, 1990), outcome considerations in blame are balanced by mental state considerations (Nelson-Le Gall, 1985). And as further differentiation of the intentional concept emerges (Baird & Malle, 2001), the distinction between justified and unjustified violations and between preventable and unpredictable outcomes emerge as well (Fincham, 1982; Shaw & Sulzer, 1964).

**What Data Would Falsify Our Proposal?**

The strongest evidence against our proposal would show that early moral evaluations or emotions in response to norm violations precede and swamp subsequent social cognitive processing (Alicke, 2000; Knobe, 2010), a reversal of what our framework suggests. Confirmation of this claim requires methods for assessing temporal and causal relations between processes (millisecond by millisecond), but such methods have yet to be introduced into moral psychology. Furthermore, confirmation of this claim requires measuring a perceivers’ affective responses after the perceivers recognize an event as norm violating but before the perceivers determine the agent’s causal involvement, intentionality, mental states, and so forth. Given the evidence for very early and automatic processing of agency and intentionality (Barrett, Todd, Miller, & Blythe, 2005; Decety, Milchalska, & Kinzler, 2012), it would be difficult, both theoretically and experimentally, to fit any kind of graded affect into this tight early time window. Nevertheless, people are likely to perceive some kind of preconceptual badness before they process all the details of a norm-violating event. Arguably, such an undifferentiated sense of badness does not represent a moral judgment (e.g., of blame), so arriving at such a judgment would require additional social cognitive processing. If this processing were systematically biased in favor of confirming the initial negative as-
A second major challenge to our proposal would be that social regulation of norm enforcement does not, as we propose, push social (and moral) cognition toward systematic information processing and accuracy. Evidence would have to show that the demand for warrant of moral judgments can be easily satisfied by biased and inaccurate social cognitive information. It would not be enough to show that under some circumstances demands for warrant are ineffective but, rather, that widespread demand for warrant either does not exist or, even if it exists as a social practice, does not predict quality of social and moral processing.

**Extension and Expansion**

**Social Regulation of Moral Judgments**

Our hypothesis that social regulation is not only the expression of moral judgment but a mechanism that keeps moral judgments honest has yet to be tested. A first requirement of testing it will be to design suitable experimental manipulations of community members putting demands on warrant for a perceiver who is expressing a moral judgment. A second requirement will be to devise reliable measures of accuracy in moral judgment and the perceiver's systematic responsiveness to evidence.

Our theoretical model predicts that the impact of demands for warrant varies by moral judgment type. Permissibility judgments are primarily reflections of shared norms, so the presence of a community member should simply increase reliability and collective agreement in these kinds of judgments, whereas the more complex blame judgments should become more deliberative and evidence-based (taking into account intentionality, mental states, etc.) in the presence of a community representative. There is also a reverse prediction—that an overwhelming need to be accepted by one's community can lead to more biased information processing if the community has strong expectations (e.g., about the guilt or innocence of a norm violator or about the appropriate level of punishment). The fine balance between these different forces may be examined with agent-based modeling methods (Elsenhörn & Gilbert, 2014). "Societies" that balance socially demanded accuracy against socially demanded unanimity should be most successful because they keep the costs of false accusations and exaggerated punishment in check. However, stratified societies in which some subgroups have more power may shift these costs to the less powerful groups. The current incarceration rates of minorities in the United States is an example of such a dynamic. As a counterforce, however, public scrutiny of aggressive policing of minorities signals a renewed demand for warrant for social–moral blame and punishment.

Institutional mechanisms of regulation, such as the state and the law, were long believed to be the dominant forms of regulation. But evidence from the fields of anthropology, psychology, sociology, and legal studies suggests that informal, interpersonal moral regulation is evolutionarily and culturally old, arising developmentally early, and is the predominant way, even today, of keeping individual community members in line with collective interests. Referring back to our flow diagram (Figure 12.3), ordinary social regulation sometimes fails; an enticing research direction might be to examine when institutional mechanisms take over social regulation and when these mechanisms are more effective than interpersonal ones.

**Affect and Emotion as Social–Moral Signals**

Although the exact causal roles of affect and emotion in the information-processing phase of moral cognition are still under debate, their involvement in public expressions of moral criticism may be more readily apparent (Wolf, 2011). Affect intensity—in words, face, and posture—seems such expressions (Voilkis et al., 2014) so that others recognize one's degree of outrage (McGeer, 2012; de Melo, Carnevale, Read, & Gratch, 2014). These expressions signal how important the violated norm is to the blamer, teach young community members about such up-importance rankings, and also communicate to norm violators what possible other sanctions might follow if they show no insight or atonement. Evidence for this social function of moral emotions might come from physiological studies that show a ramping up of negative arousal from early violation detection to late public expression. That is, the very opportunity to express one's judgment publicly may increase the importance of affect that was during mere "in the head" judgments, quite modest. Additional support might come from evidence that perceivers have less differentiated emotions when they directly form their moral judgments than when they publicly express them, because anticipatory public scrutiny leads to more attentive information appraisals. Here, too, perceivers' perception of a community's strong expectations may sometimes unduly modulate their public judgments, such as offering exaggerated expressions of outrage; this, in turn, can fuel even stronger expressions by other community members and escalate collective moral condemnation beyond what perceivers felt in private.

**Artificial Morality**

Work on moral psychology has recently expanded into artificial morality—the study and design of computational models of moral competence (Mao & Gratch, 2012; Tomat & Forbus, 2008) and implementation in social robots (Wallach & Allen, 2008; Malle & Scheutz, 2014). Social robots—embodied machines that are able to interact with humans—play an increasing role in contemporary society. Around a decade ago there were no robots in private homes, whereas in 2014, 4.7 million service robots for personal and domestic use were sold worldwide (International Federation of Robotics, 2015). These robots rarely possess extensive social cognitive capacities but are improving rapidly (Nourbakhsh, 2015), and robots may soon function as social companions or assistants in health care, education, security, and emergency response. In such applications, however, robots will need to have basic moral competence to ensure physically and psychologically safe interactions with humans (Malle & Scheutz, 2014). Designing such robots offers appealing new avenues for research, by testing more precise, formally specified models of social interaction (e.g., making interpolation inferences in live interactions) and moral capacities (e.g., recognizing norm violations and forming-based judgments). In addition, research will need to identify the conditions under which humans ascribe features such as intentionality, free will, or basic social knowledge to artificial agents (Malle et al., 2015; Meltzoff, Brooks, Shon, & Rao, 2010; Monroe, Dillon, & Malle, 2014), because such ascriptions fundamentally alter human–machine interactions. Integrating robots into society will require such understanding to achieve beneficial human–robot coexistence.

**Summary**

Returning from social cognitive science fiction, we close by recapping our theoretical framework for understanding the processes of moral cognition. We argue that a hierarchy of social cognitive tools grounds moral cognition and that social and moral cognition together guide the social regulation of behavior. The practice of social–moral regulation, in turn, puts pressure on community members to engage in reasonably fair and evidence-based moral criticism. With the help of these cognitive adaptations and social practices, people are able to navigate the terrain of morality, accruing bumps and bruises along the way but surviving as the most sophisticated social creature currently roaming the earth.

**ACKNOWLEDGMENTS**

We are grateful for the collective work and insights by all members of our lab ([http://research.clps.brown.edu/SocCogSci/Personnel/personnel.html](http://research.clps.brown.edu/SocCogSci/Personnel/personnel.html)). This project was supported by grants from the Office of Naval Research (Nos. N00014-14-1-0144 and N00014-13-1-0269).

**NOTES**

1. Perhaps a helpful term for this set of capacities would be moral competence (Malle, 2016; Malle & Scheutz, 2014). A complete rendering of this competence would include both moral and tentionality inferences in live interactions) and moral capacities (e.g., recognizing norm violations and forming-based judgments). In addition, research will need to identify the conditions under which humans ascribe features such as intentionality, free will, or basic social knowledge to artificial agents (Malle et al., 2015; Meltzoff, Brooks, Shon, & Rao, 2010; Monroe, Dillon, & Malle, 2014), because such ascriptions fundamentally alter human–machine interactions. Integrating robots into society will require such understanding to achieve beneficial human–robot coexistence.
positive and negative behaviors, but here we focus, in keeping with the literature, on nega-
tive behaviors.

2. These socially corrective strategies are not in-
ventions of modern legal institutions; rather, they are successful informal practices that have persisted throughout history (Boehm, 1999; Pospisil, 1971).

3. There are well-known limits to this shaping process: For example, members of a given group may demand fair and accurate norm enforcement for one another but not for mem-
bers of disliked or lower status outgroups.

REFERENCES


Moral Cognition and its Basis

119

...


