Annual Review of Developmental Psychology

Human Morality Is Based on an Early-Emerging Moral Core

Brandon M. Woo,[1] Enda Tan,[2] and J. Kiley Hamlin[3]

[1]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA; email: bmwoo@g.harvard.edu

[2]Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland, USA; email: endatan@umd.edu

[3]Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada; email: kiley.hamlin@psych.ubc.ca

Annu. Rev. Dev. Psychol. 2022. 4:7.1–7.21

The *Annual Review of Developmental Psychology* is online at devpsych.annualreviews.org

https://doi.org/10.1146/annurev-devpsych-121020-023312

Keywords

social evaluation, social cognitive development, morality, infancy

Abstract

Scholars from across the social sciences, biological sciences, and humanities have long emphasized the role of human morality in supporting cooperation. How does morality arise in human development? One possibility is that morality is acquired through years of socialization and active learning. Alternatively, morality is instead based on a "moral core": primitive abilities that emerge in infancy to make sense of morally relevant behaviors. Here, we review evidence that infants and toddlers understand a variety of morally relevant behaviors and readily evaluate agents who engage in them. These abilities appear to be rooted in the goals and intentions driving agents' morally relevant behaviors and are sensitive to group membership. This evidence is consistent with a moral core, which may support later social and moral development and ultimately be leveraged for human cooperation.

Contents

INTRODUCTION

What makes humans special? Darwin (1871) once wrote that "of all the differences between [humans] and [nonhuman] animals, the moral sense or conscience is by far the most important" (p. 67). Indeed, humans readily help others and avoid harming them, even at a cost to their own fitness. Further, they show concern for how people treat each other and negatively evaluate and punish those who treat others poorly. Remarkably, humans do these things not only when they have some connection to the agents involved or otherwise stand to directly gain or lose from doing so but also when they are entirely uninvolved bystanders (Fehr & Fischbacher 2003, Fehr & Gächter 2002, Greene & Haidt 2002). Psychological and anthropological research has documented high levels of similarity in which concepts are deemed morally relevant across cultures and what behaviors are deemed morally right (e.g., helping, being fair, respecting superiors) versus wrong (Brown 1991, Curry et al. 2019, Graham et al. 2011; see also Atari et al. 2022). These findings point to a largely shared moral sense in humans. What are the origins of this moral sense?

Traditionally, developmental psychologists have examined humans' moral origins by examining how children's experiences in and with the world facilitate the emergence and development of a moral sense. For instance, senior members of society teach or otherwise socialize junior members about right and wrong; junior members eventually internalize these lessons as their own (Bandura 2014, Dunn 2014, Grusec et al. 2014, Kuczynski & Knafo 2014). Children also take active roles in their moral development, reflecting on and interpreting morally relevant experiences (Piaget

1932; Smetana 2006; Smetana et al. 2014; Turiel 1983, 2014). For instance, in response to toddlers' moral transgressions, caregivers discuss them with their toddlers and encourage thinking about the consequences of those transgressions for others (Dahl 2016, Dahl & Campos 2013, Smetana 1989). Studies have linked such practices to variation in moral trajectories (see, e.g., Augustine & Stifter 2015, Kochanska 2002, Smetana et al. 2012, Yoo & Smetana 2021), highlighting how both socialization and active reflection support developing a moral sense.

Outside of developmental psychology, scholars from diverse disciplines including evolutionary biology, anthropology, economics, political science, and philosophy have focused on morality's evolutionary origins, stressing the role of the moral sense in supporting cooperation (Alexander 2007, Cosmides & Tooby 1992, Curry et al. 2019, Darwin 1871, Gintis et al. 2008, Henrich & Henrich 2007, Joyce 2007, Katz 2000, Sterelny et al. 2013). Indeed, cooperation stands at the foundation of group living in diverse species, with humans cooperating with nonkin at levels unmatched in the animal kingdom (Axelrod 1984, Trivers 1971). Critically, cooperative systems are at risk from those who take advantage of others' cooperative acts without reciprocating. Here, the sense that some actions are permissible/good/right and others are impermissible/bad/wrong serves to discourage free riding, creating stability in an otherwise unstable system. Thus, humans' moral sense may have evolved alongside their cooperative tendencies, allowing cooperative systems to flourish.

Of course, the ultimate purpose of the moral sense need not impact its emergence during individual life spans. Very young children arguably have little use for a moral sense, assuming their caregivers possess it. That said, studies of infants and toddlers have revealed abilities that emerge surprisingly early in various other key cognitive domains, including objects, numbers, space, and agents; some interpret these abilities as evidence that humans enter the world with core knowledge, or skeletal systems for understanding and reasoning about a few key domains that support rapid additional learning (Carey 2009, Spelke & Kinzler 2007). Further, humans depend on others from birth to help them achieve their goals (Tomasello & Carpenter 2007), teach them new knowledge and skills (Gweon 2021), and provide them with food and shelter (Hrdy 2009); indeed, in many human societies both parents and nonparents engage in caregiving (Hawkes 2014; Hrdy 1999, 2009). This high degree of interdependence among nonkin raises the possibility that even infants could benefit from abilities to assess others' cooperative potential. Given this, humans' systems of core knowledge may include primitive abilities to understand and evaluate morally relevant actions, ones that do not rely wholly on socialization and learning (Hamlin 2013b, Rottman & Young 2015; see also Ting et al. 2019). We term these abilities a "moral core."

In the past two decades, developmental psychologists have provided crucial tests of the possibility that there is a moral core by studying sensitivity to the moral world in infants and toddlers. These tests are based on the premise that very young children are incapable, or at least far less capable, of internalizing or actively constructing moral norms relative to the older children that are typically examined in research on socialization and active learning processes. To the extent that infants and toddlers understand and evaluate morally relevant behavior, then, this would support the possibility that a moral core exists. Below, we review evidence that infants and toddlers understand and evaluate morally relevant behaviors performed by novel third parties. Some evidence suggests that basic forms of these abilities are present by just 3 months after birth, with abilities reaching rather striking levels of sophistication before the end of the first year and rapidly progressing thereafter. The early emergence and rapid expansion of these capacities are not easily explained by socialization and active learning; they instead suggest that processes of socialization and active learning serve to facilitate the expansion of a moral core rather than build it from scratch.

EXPLORING INFANTS' AND TODDLERS' ASSESSMENTS OF MORALLY RELEVANT BEHAVIOR

How could one examine whether infants and toddlers understand and evaluate morally relevant behavior? Whereas adults and older children can understand verbal vignettes and answer questions about them, infants and younger toddlers cannot. Thus, researchers wishing to examine infants' moral capacities are restricted to presenting visual depictions of morally relevant acts and examining nonverbal responses (e.g., looking, reaching, giving); from these responses, understanding and evaluation must be inferred (cf. Tafreshi et al. 2014). To illustrate, imagine an agent (a "protagonist") who tries but fails to climb a hill (**Figure 1*a***). Sometimes, a second agent pushes the protagonist up the hill, allowing it to achieve its goal. Other times, a third agent pushes the protagonist down the hill, preventing it from achieving its goal. As adults, we describe the second agent's

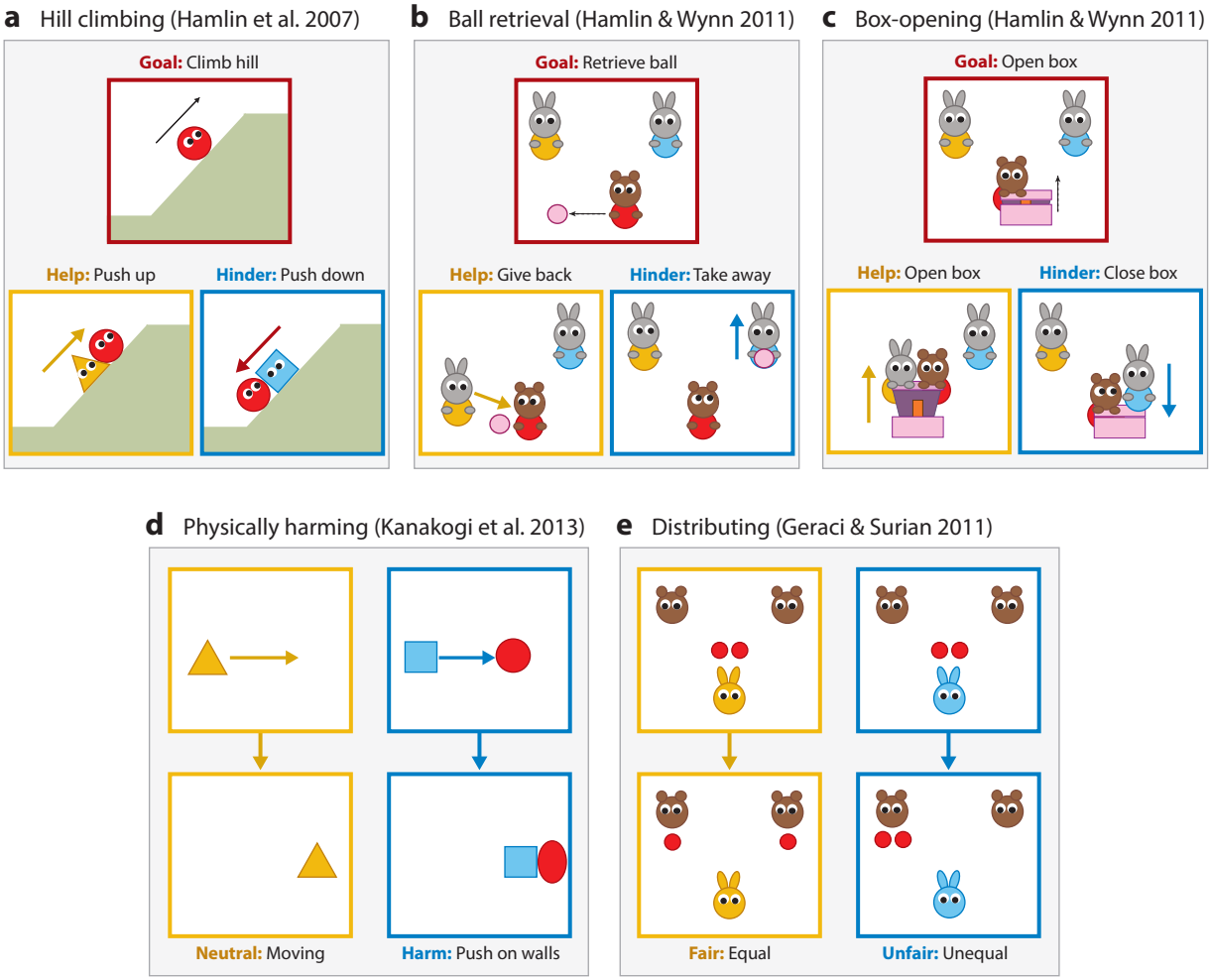

Figure 1

Schematics of some of the scenarios used to probe early assessments of morally relevant behavior. Scenarios have depicted (*a–c*) helping and hindering, (*d*) physical harm, and (*e*) fair and unfair distributions. Note that agents in the displays of Kanakogi et al. (2013) lacked eyes.

actions as "helping the protagonist" and the third agent as "hindering the protagonist." Moreover, we find the second agent nicer, prefer the helper over the hinderer, and expect the protagonist to feel the same. Do infants make these inferences and form these evaluations?

To probe this question, Hamlin et al. (2007) presented 6- and 10-month-olds with the hill-climbing events described above, wherein each agent was a different colored wooden shape with eyes. Infants were shown helping and hindering events until habituating (Colombo & Mitchell 2009) to give them sufficient opportunity to process the interactions. Subsequently, an experimenter—unaware of each agent's prior actions—placed the helper and the hinderer within the infants' reach and observed which agent infants touched first. Here, both 6- and 10-month-olds preferentially reached for the helper over the hinderer. In subsequent experiments involving either a helper or a hinderer and a neutral agent, infants preferred a helper over a neutral agent, and a neutral agent over a hinderer. Together with studies replicating these patterns (Hamlin 2015, Loheide-Niesmann et al. 2021; cf. Schlingloff et al. 2020) these results suggested that infants (at least within this single prosocial/antisocial interaction) differentiated helping from hindering, positively evaluated the helper, and negatively evaluated the hinderer.

But do these abilities reflect a moral core? There are many ways to approach this question, which we return to throughout this review. First, given that core cognitive capacities are theorized to not be rooted in experience, one might consider whether infants have had opportunities to learn these things. Although it is difficult to imagine attempting to socialize 6-month-olds as to the value of helping versus hindering, by this age infants have likely had many experiences that could support such evaluations. Indeed, although they cannot self-locomote, 6-month-olds have presumably unsuccessfully pursued other goals, been helped or hindered, and felt pleasure or displeasure as a result. They may also have observed others experiencing the same. If so, perhaps infants extend these experiences onto helping and hindering among third parties. Although it is difficult to understand such experiences without measuring them directly (a key avenue for future work), studies of even younger infants, with more severe physical limitations and less observational experience, provide stronger tests of the possibility that there is a moral core.

Thus, Hamlin et al. (2010) examined 3-month-olds' preferences for the helper versus the hinderer from the hill-climbing scenario (via preferential looking, as 3-month-olds are not yet able to reach for objects). Here, 3-month-olds preferred looking at the helper over the hinderer, mirroring older infants' reaching, and the neutral agent over the hinderer. Interestingly, unlike older infants, 3-month-olds looked equally toward the helper and the neutral agent, suggesting that negative evaluations of hindering may emerge prior to positive evaluations of helping (for a review of negativity biases in infancy, see Vaish et al. 2008). Taken together with Hamlin et al. (2007), these findings provide evidence that infants understand and differentially evaluate actions that help versus hinder an agent's unfulfilled goal, even when their own relevant experience is arguably extremely limited.

Notably, although infants in these studies appeared to differentiate helping and hindering on some basis (as otherwise, they should not have been able to differentially evaluate helpers versus hinderers), they did not appear to expect novel agents to engage in helping versus hindering. A long tradition of infancy research suggests that infants look longer at events that are surprising or unexpected (Aslin 2007). In Hamlin and colleagues' (2007, 2010) studies, infants consistently looked equally long after helping and hindering events, as though they found them equally likely to occur (for evidence that providing 17-month-olds with stronger evidence of a protagonist's goal leads them to expect agents to help the protagonist, cf. Lee et al. 2020). Yet it was not that infants held no expectations with respect to these events: 10-month-olds (though not younger infants) looked reliably longer when the protagonist approached the hinderer versus the helper, as though

they expected the protagonist to also prefer the helper (see also Fawcett & Liszkowski 2012, Lee et al. 2015; cf. Kuhlmeier et al. 2003).

In sum, studies using the hill paradigm suggest that infants evaluate agents who help and hinder and, by the end of the first year, expect others to share those evaluations. These tendencies have been demonstrated by just 3 months of age, rendering it less likely—though not impossible—that experience is solely responsible for their emergence. These abilities, however, need not reflect a moral core. Within the hill paradigm (or any given pair of prosocial/antisocial events), various features distinguish helping from hindering, many without moral, or even social, content. For instance, perhaps infants' responses were based on physical differences between helping and hindering (see, e.g., Scarf et al. 2012a,b; cf. Hamlin 2015). Indeed, unless helping and hindering acts were somehow physically identical, it would be difficult to rule out this possibility within a single paradigm.

To date, studies have addressed the possibility that infants and toddlers are responding to the physical features of helping and hindering in two ways; these are reviewed below. First, studies have examined the consistency of infants' and toddlers' responses across physically distinct instances of helping and hindering and among prosocial and antisocial acts more broadly. If infants and toddlers respond similarly across various physical instantiations of prosocial and antisocial acts, this would suggest that their responses are not physical in nature. Second, studies have examined whether infants and toddlers respond to the same physical features that distinguish prosocial from antisocial acts in situations that are not social and so are morally irrelevant. Indeed, notions of prosocial and antisocial behavior critically involve social interactions in which agents' actions are directed at other agents; the same physical behaviors that help and hinder an animate agent are typically neither prosocial nor antisocial if directed at something inanimate. Thus, various inanimate control conditions have shown infants events in which the very same agents perform the very same acts but directed toward inanimate objects. If infants and toddlers demonstrate evaluations or expectations in these control conditions, this suggests that their responses were not based solely on physical features in the social conditions.

Distinct Acts of Helping and Hindering

To date, studies have presented infants with protagonists being helped and hindered while pursuing various unfulfilled goals. For instance, in the ball scenario (Hamlin & Wynn 2011), an animal-puppet protagonist loses control of a ball (**Figure 1*b***). The helper (another puppet, wearing a colored t-shirt) returns the ball, whereas the hinderer (a third puppet, wearing a different t-shirt) takes the ball away. In this study, both 5- and 3-month-olds preferentially reached for or looked at the helper, respectively, over the hinderer, as infants had in the hill scenario (see also Holvoet et al. 2018, Scola et al. 2015). Notably, infants did not distinguish between agents who gave to versus took from an inanimate claw, as though they were responding to the social, versus the physical, nature of giving and taking. Likewise, in the box scenario (Hamlin & Wynn 2011), an animal-puppet protagonist unsuccessfully tries to open a box containing a toy (**Figure 1*c***). The helper opens the box, allowing the protagonist to grasp the toy; the hinderer slams the lid shut, preventing access. Here, both 5- and 9-month-olds reached for the helper over the hinderer (cf. Salvadori et al. 2015); they did not do so when actions were directed toward an inanimate claw. Additional studies have shown older toddlers these and other instances of helping and hindering (e.g., helping an agent get a desired toy off a shelf versus stealing the toy) and found that toddlers are more prosocial toward helpers than hinderers (e.g., giving helpers toys) (Hamlin et al. 2011, Van de Vondervoort et al. 2018).

Interestingly, in these new scenarios too, infants and toddlers have not reliably looked longer after events in which agents hinder versus help other agents. In addition to suggesting that infants

and toddlers likely do not hold default expectations that agents will help versus hinder others, they suggest that a tendency to pay differential attention following helping or hindering is unlikely to explain infants' and toddlers' preferences for helpers.

Physical Harm

Across cultures, adults consider harm to be morally relevant (Curry et al. 2019, Graham et al. 2011). To probe how infants respond to agents who physically harm others, Kanakogi et al. (2013) presented 10-month-olds with one animated-shape agent who acted aggressively toward a victim agent (i.e., running into it) and another (neutral) agent who engaged in similar movements directed at empty space (**Figure 1*d***). Infants preferentially reached for the neutral agent, as though they negatively evaluated the aggressor. Converging evidence has come from Buon et al. (2014); here, 10-month-olds preferentially accepted toys from a human who comforted another human (and kicked an inanimate backpack) over one who kicked another human (and comforted a backpack). Across these studies, then, infants avoided agents who physically harmed other agents.

If infants negatively evaluate physical harm, they may also positively evaluate agents who prevent it. Kanakogi et al. (2017) presented 6-month-olds with aggressive hitting events in which one bystander intervened by inserting itself between the aggressor and victim, preventing harm; another bystander moved elsewhere, failing to prevent harm. Here, 6-month-olds reached for the agent who prevented harm over the agent who did not. These results have been conceptually replicated in 17-, 20-, and 21-month-olds; importantly, toddlers do not form preferences when hitting is directed at inanimate objects (Geraci 2020a,b; Geraci & Simion 2021). Across studies, then, infants and toddlers selectively approached agents who prevented harm to other agents.

Explorations of infants' and toddlers' attention following harming and harm prevention have revealed similar results to those following helping and hindering. There is evidence that infants expect others to positively evaluate those who prevent harm (as they do themselves). In one study, 20-month-olds looked longer when bystanders approached noninterveners versus interveners (Geraci 2020b). However, there is no evidence either that infants find physical harm surprising or that they expect bystanders to prevent harm. Studies reporting looking times following these events have revealed no systematic attentional differences (e.g., Kanakogi et al. 2013, 2017). In sum, then, infants appear to understand and evaluate actions within the domain of physical harm: They negatively evaluate agents who harm others and positively evaluate agents who prevent harm.

Fairness

Adults tend to view equal resource distributions as fair and unequal resource distributions as unfair. Absent contextual information, we more positively evaluate fair versus unfair distributors (Curry et al. 2019). In a study probing early responses to resource distributions, Schmidt & Sommerville (2011) presented 15-month-olds with an agent who distributed resources either equally or unequally among two recipients (e.g., splitting 4 crackers 2:2 or 3:1). Here, 15-month-olds looked longer following the unequal versus the equal distribution; they did not look differently when there were no recipients present. Converging evidence for fairness expectations has come from wide age ranges (from 4 to 21 months), following distributions of a variety of resources (e.g., milk, toys; Buyukozer Dawkins et al. 2019, Enright et al. 2017, Meristo et al. 2016, Sloane et al. 2012, Sommerville et al. 2013, Ziv & Sommerville 2017), and done by various types of agents (humans, puppets, cartoon agents). Although there have been null findings in younger infants (e.g., Sommerville et al. 2013), they appear to be explained by difficulties that younger infants have with tracking larger numbers (Buyukozer Dawkins et al. 2019). Further, inanimate control conditions

suggest that infants' fairness expectations are selective to distributions involving animate recipients (e.g., Sloane et al. 2012). This larger body of research suggests that infants and toddlers have expectations for fairness: that agents will distribute resources equally between animate recipients.

Notably, these results stand in contrast to the apparent lack of early expectations that agents will help versus hinder or that they will refrain from and/or prevent physical harm. Although the reason for this asymmetry is unclear, it could indicate that notions of fairness and help or harm are structured differently within the moral core (for further discussion, see Ting et al. 2019). Consistent with this idea, research on infants' evaluations suggests differential developmental trajectories between the fairness and helping/hindering/harm domains. Indeed, although there is evidence that infants and toddlers evaluate fair versus unfair distributors, most positive evidence comes from toddlers in the second year, who are considerably older than the 3-to-6-month-olds showing preferences within the help/hinder/harm domains. For instance, Geraci & Surian (2011) presented 10- and 16-month-olds with cartoon animals who distributed toys to two recipients either equally or unequally (**Figure 1*e***). Here, only the 16-month-olds preferentially reached for the fair over the unfair distributor—and only when the recipients were animate. In other studies, 13-, 15-, and 17-month-olds have preferentially approached and accepted toys from fair versus unfair humans (Burns & Sommerville 2014, Lucca et al. 2018). Likewise, in Ziv et al. (2021), 16-month-olds were more likely to give resources to than remove resources from a fair distributor but not an unfair distributor, and in DesChamps et al. (2016), 13- and 15-month-olds expected fair distributors to be praised and unfair distributors to be admonished. Thus, infants and toddlers do positively evaluate fair distributors but perhaps not until the second year of life (cf. Geraci et al. 2022).

Interim Summary

In the above, we have reviewed evidence that infants and toddlers understand and evaluate multiple morally relevant behaviors: They more positively evaluate helpers, protectors, and fair distributors than they do hinderers, nonprotectors, and unfair distributors, and they expect others to do the same. Although there is evidence for evaluations within the help/hinder/harm domains earlier than in the fairness domain, neither infants nor toddlers appear to hold baseline expectations that agents will help and not harm each other, whereas even young infants seem to expect fair rather than unfair distributions. These responses appear not to stem from physical features of events; there are significant physical differences between each type of prosocial/antisocial behavior, and although across studies agents have been represented by humans, puppets, and cartoons, infants and toddlers do not distinguish between the same acts directed toward inanimate objects. Thus, infants' and toddlers' responses are consistent with the existence of a general capacity to understand and evaluate morally relevant behavior.

Replicability concerns. Before reviewing evidence that infants and toddlers understand and evaluate more complex morally relevant behaviors, it is important to address the issue of replicability. Indeed, to date there have been several failures to replicate infants' preferences for helpers versus hinderers (e.g., Cowell & Decety 2015, Salvadori et al. 2015, Schlingloff et al. 2020, Vaporova & Zmyj 2020); these failures have inspired questions about the robustness of early responses to morally relevant events. Because the possibility that infants possess a moral core requires these results to be robust, we discuss here our view of the (current) state of the evidence that infants prefer prosocial to antisocial others.

Margoni & Surian (2018) gathered all published and unpublished data in a meta-analysis of infants' and toddlers' evaluations in studies of helping and hindering, physical harm, and fair and unfair distributions. Across studies and dependent variables (reaching, giving, or helping by

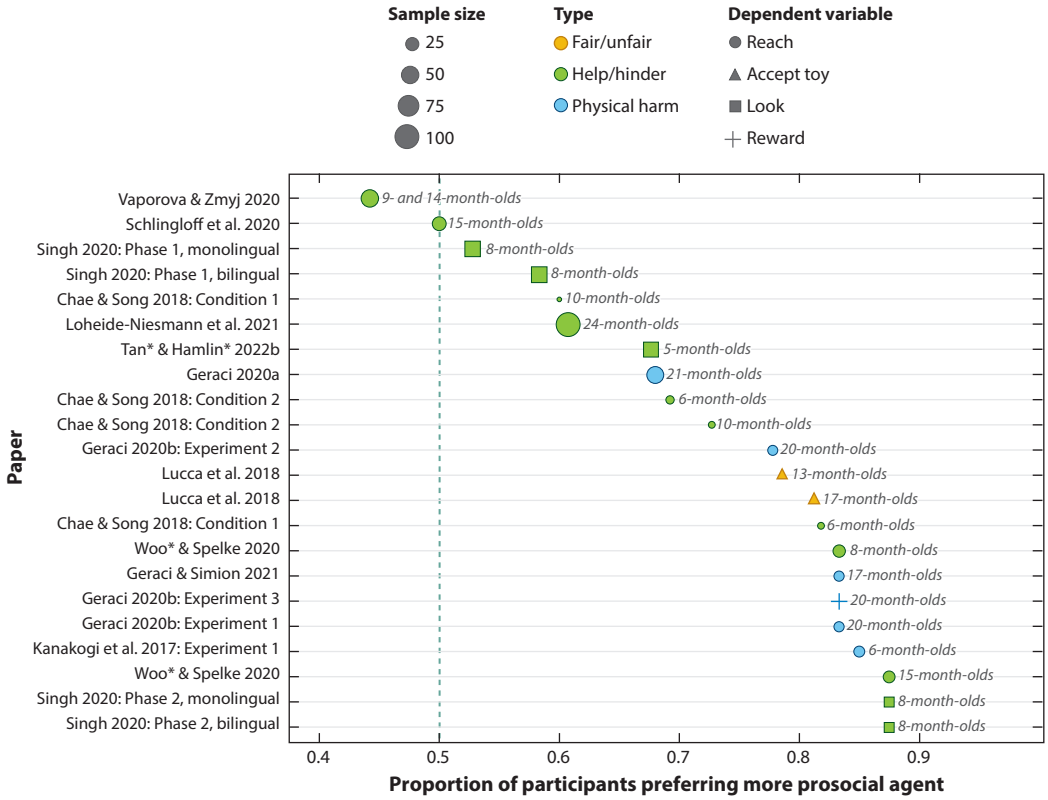

Figure 2

Results of new experiments (since Margoni & Surian 2018) probing infants' and toddlers' evaluations of prosocial, antisocial, or neutral agents. This graph includes only experiments in which agents' intentions were consistent with the outcomes they caused and actions were prosocial. The dependent variable was either necessarily binomial (e.g., reaching or giving behaviors) or binarized (e.g., which agent infants looked to longer for preferential looking DVs) so that data could be compared. The size of the circles indicates sample size, color indicates the type of morally relevant behavior that a task involved, shape indicates the type of DV, asterisks indicate an author of the present review, and the dashed line indicates chance responding (i.e., 50%).

participants), about two-thirds of infants and toddlers preferred prosocial to antisocial characters. Following this meta-analysis, at least 15 additional papers probing early evaluations have been published: 13 report evidence that infants and toddlers prefer prosocial agents and/or avoid antisocial ones in at least one condition (Chae & Song 2018; Geraci 2020a,b; Geraci & Simion 2021; Holvoet et al. 2018; Kanakogi et al. 2017; Loheide-Niesmann et al. 2021; Lucca et al. 2018; Singh 2020; Tan & Hamlin 2022b; Van de Vondervoort et al. 2018; Woo & Spelke 2020; Ziv et al. 2021), and three report null findings (Schlingloff et al. 2020, Singh 2020, Vaporova & Zmyj 2020) (**Figure 2**). These studies, combined with Margoni & Surian's (2018) meta-analysis, provide an overall picture that infants and toddlers understand and evaluate prosocial and antisocial agents.

THE BASIS OF INFANTS' AND TODDLERS' ASSESSMENTS OF MORALLY RELEVANT BEHAVIOR

Even if these results are robust and selective to situations involving social interactions, the question remains as to the basis on which infants and toddlers distinguish morally relevant behaviors. Adults may label these behaviors as helping and hindering, fair and unfair, protective and aggressive, and,

more broadly, prosocial and antisocial; these terms have clear moral implications. Yet infants and toddlers may not distinguish these behaviors in moral ways. In what follows, we consider various more and less moral ways in which infants and toddlers could interpret prosocial or antisocial acts and review evidence pitting distinct interpretations against each other. We argue that the most parsimonious explanation for the extant evidence is that infants and toddlers are sensitive to the morally relevant features of prosocial and antisocial acts.

Nonmoral Social Accounts of Infants' and Toddlers' Assessments

Are infants and toddlers sensitive to morally relevant behavior, or are they picking up on other, nonmoral social cues? To date, various researchers have pointed out that infants may prefer helpers to hinderers not because they facilitate versus prevent another agent's goal (that is, not because they are or are not helpful), but because they are more likely to act like or move concordantly with a protagonist (Benton & Lapan 2022, Powell 2019, Powell & Spelke 2018). Although these two theories vary somewhat, as do their explanations for where infants' assessments originate from, they both hold that infants are sensitive to how social agents are, prefer agents who are more social, and expect others to do the same. On both accounts, the sociality of an agent can be inferred from their movement. To illustrate, in the hill scenario, after the protagonist attempts to climb a hill, the helper moves with and in the same direction as the protagonist; the hinderer, by contrast, moves against and in the opposite direction. Thus, the helper may appear more social than the hinderer: They engage in concordant as opposed to discordant movement with another agent and move like another agent rather than in a different way. Indeed, Powell & Spelke (2018) have found that infants prefer those who imitate others' actions over those who do not, including cases in which agents imitate movement up a hill.

Although we agree that some studies are clearly susceptible to these movement-based alternative nonmoral social explanations, in our view they fail to account for the broader set of findings. These explanations do not easily account for evaluations of agents who prevent versus cause physical harm, nor of agents who engage in fair versus unfair distributions; in these scenarios the more prosocial agents neither imitate nor move concordantly with others. Moreover, there is growing evidence that infants' and toddlers' assessments of prosocial and antisocial acts rely—perhaps exclusively—on the mental states that drive prosocial and antisocial acts. This body of evidence renders movement-based social explanations for infants' responses less plausible.

The Role of Mental States in Infants' and Toddlers' Assessments

Representations of agents' mental states are central to adults' moral reasoning. For instance, to understand how actions help versus hinder an agent in pursuit of a goal, one must first represent that agent's goal (e.g., to climb a hill). Likewise, to accurately assess an agent's cooperative potential, one needs to track not only the outcomes of others' actions (e.g., that a helper's actions led to the protagonist reaching the top of the hill and a hinderer's actions did not), but also the prosocial and antisocial intentions underlying others' actions (e.g., that one agent wanted to help and the other wanted to hinder the protagonist). Indeed, it is presumably these intentions that are diagnostic of future cooperative acts (Heider 1958, Malle 1999). Are infants and toddlers sensitive to the mental states underlying others' actions, or are their assessments more superficial?

The goal of an agent in need of help. Here, we highlight three studies that speak to whether early evaluations of helping and hindering reflect an understanding of the goal of an agent in need of help. First, Hamlin (2015) adapted the hill-climbing scenario of Hamlin et al. (2007), but with one critical change: The protagonist's googly eyes had unfixed pupils, so that as it moved up the

hill, it appeared to look downward (see also Scarf et al. 2012b). Infants are highly sensitive to gaze direction (e.g., Hood et al. 1998); without looking up the hill, the protagonist's actions should not appear goal-directed. Indeed, 6-to-11-month-olds preferred an agent who pushed the protagonist up the hill over an agent who pushed the protagonist down the hill only when the protagonist's pupils were fixed toward the top of the hill, reflecting its goal; when the pupils were unfixed, infants did not distinguish the agents. These findings suggest that clear goal information is necessary for infants' evaluations of helpers and hinderers.

Second, using eye-tracking, Tan & Hamlin (2022b) examined where 5-month-olds looked when watching cartoon versions of the hill-climbing scenario. Here, the more infants looked to the target of a protagonist's goal (the top of the hill), the stronger infants later preferred looking to the helper over the hinderer. These findings provide further support for the role of goal understanding in infants' evaluations of helpers and hinderers.

Third, Woo & Spelke (2022b) examined how infants' and toddlers' developing action understanding directly informs their evaluations of helping. Developmental studies of action understanding have revealed that whereas toddlers and older infants can infer the ultimate goal of means–end actions (e.g., opening a box in order to grasp a toy inside), younger infants are more focused on means states (here, opening the box itself) (e.g., Sommerville & Woodward 2005). Capitalizing on this developmental change, Woo & Spelke (2022b) found that infants and toddlers evaluated those performing the very same box-opening actions differently depending on whether their age made them more likely to see a protagonist who struggled to open a box as having the goal of getting the toy inside the box (15-month-olds) versus opening the box itself (8-month-olds). When the location of the contained object changed to a different box, 8-month-olds preferred a puppet who opened the old box for the protagonist, whereas 15-month-olds preferred a puppet who opened the new box for the protagonist.

In sum, then, growing evidence demonstrates a direct link between infants' and toddlers' goal understanding to their evaluations of helping and hindering. Notably, in these studies, infants' distinct responses do not rely on changes in agents' respective movement patterns but only on changes in implied and/or inferred mental states; this provides further evidence that infants' and toddlers' responses to morally relevant behaviors are not attributable to physical features of events.

The intentions of agents who engage in morally relevant actions. In the studies reviewed so far, prosocial and antisocial agents have successfully carried out their prosocial and antisocial intentions. That is, the outcomes that agents cause for others are always consistent with their intentions (e.g., when an agent tries to act fairly, the recipient is treated fairly). Given this, infants' and toddlers' assessments of morally relevant behaviors might be based on prosocial and antisocial intentions, positive and negative outcomes, or both. Adults are sensitive to others' intentions when reasoning about moral actions (Cushman et al. 2006), generally privileging intention over outcome in their moral judgments (for cross-cultural differences, cf. McNamara et al. 2019; see also Saxe 2016). Are infants' and toddlers' assessments based on others' prosocial and antisocial intentions, or are they more shallow, focused on the outcomes that agents cause or are associated with? Here, we highlight two bodies of evidence that suggest that infants' and toddlers' understanding and evaluations of prosocial and antisocial acts are based on intentions, with evidence coming from each of the morally relevant domains reviewed.

The first body of evidence comes from research on infants' assessments of failed attempts. By late in the first year, infants and toddlers appear to track agents' attempts to be prosocial or antisocial, even if agents' attempts are unsuccessful. In the domain of helping and hindering, 8-month-olds (but not 5-month-olds) preferred an agent who tried but failed to help a protagonist open a box (the box remained closed) over an agent who tried but failed to hinder the protagonist

in opening the box (the protagonist opened the box) (Hamlin 2013a). Similarly, in the domain of harm, 10-month-olds (but not 6-month-olds) preferred an agent who actively tried to prevent harm over one who did not make such attempts but just happened to prevent harm (Kanakogi et al. 2017). Finally, 9-month-olds preferred an agent who tried but failed to distribute resources equally between two potential recipients over an agent who tried but failed to distribute resources unequally; in both cases, because the recipients were at the tops of two steep hills, and the agents struggled to climb those hills with the resources, no recipients ever received resources. Together, these results suggest that by 8–10 months of age, infants positively evaluate those who try to be prosocial, irrespective of subsequent outcomes. Further, although younger infants below ∼8 months of age have failed to show sensitivity to intention in situations in which intention and outcomes are inconsistent, even when outcomes could be used to distinguish between distinct agents, the infants do not do distinguish between the agents (Hamlin 2013a, Kanakogi et al. 2017). Thus, the evidence suggests that there is no outcome bias operating during infancy (see also Margoni & Surian 2016).

By late in the first year, infants also appear to base their expectations for how others will respond to morally relevant actions on the intentions underlying those actions, as revealed by situations involving failed attempts. In the domain of helping and hindering, 12-month-olds expected a protagonist to approach an agent who had attempted to help the protagonist up a hill over another agent who had instead attempted to hinder the protagonist, even though infants never saw the outcome of the agents' actions (Lee et al. 2015). Moreover, in the domain of fairness, following displays such as those of Geraci et al. (2022), 10-month-olds expected an uninvolved bystander to approach an agent who had attempted to be fair over another agent who had attempted to be unfair, even though outcomes were the same for all recipients. Thus, infants' and toddlers' expectations converge with their evaluations: By late in the first year, infants and toddlers appear sensitive to the intentions underlying others' morally relevant actions.

A second set of evidence suggesting that infants' and toddlers' assessments are based on intentions comes from studies examining states of knowledge and ignorance. As mature moral observers, adults appear to track what agents know versus are ignorant about and tend to hold agents responsible for only the positive and negative outcomes they knowingly (and, presumably, intentionally) cause. For example, the same act of adding peanuts to a dish can be malicious if one thinks that a friend is allergic to peanuts or unfortunate but not blameworthy if one is unaware of the allergy (Young et al. 2007). Do infants and toddlers restrict their evaluations to those who engage in actions with foreknowledge of how their actions are likely to influence other agents? How do infants and toddlers make sense of agents whose intentional actions have positive and negative consequences for other agents but do not have awareness of what those consequences would be?

To probe these questions, Hamlin et al. (2013b) presented 10-month-olds with a protagonist who persistently approached one of two toys in the presence or absence of two other agents. Following familiarization, the protagonist lost access to both toys, and one agent provided access to the desired toy (a positive outcome), whereas the other provided access to the nondesired toy. Here, rather than consistently preferring the agent who caused the positive outcome, infants instead did so only when the agents had been present as the protagonist revealed its preference for that toy. When the agents had instead been absent, infants chose randomly between the agents. These findings have been conceptually replicated and extended in additional situations involving helping and hindering (Woo et al. 2017; Woo & Spelke 2022a,c).

Taken together, these findings suggest that infants' and toddlers' understanding and evaluations of morally relevant behavior are not superficially based on outcomes but instead are based on the intentions underlying agents' actions. These studies have two key implications for the possibility of a moral core. First, as noted above, adults' moral judgments focus on others' intentions; infants'

focus on intentions, at least by the end of the first year, suggests that their responses may reflect early moral judgments. Second, in most studies cited above, infants and toddlers responded to the same action differently, solely dependent on variations in goals and intentions; thus, movement-based social explanations are unlikely to account for these responses.

THE ROLE OF GROUP MEMBERSHIP IN INFANTS' AND TODDLERS' ASSESSMENTS

The studies reviewed above have examined infants' and toddlers' evaluations of social interactions among strangers. In real life, however, adults' and children's understanding and evaluations of prosocial and antisocial acts are often biased by their social identities: The same act (e.g., harm) can be evaluated differently depending on whether it is directed toward an ingroup or outgroup member (Curry et al. 2019, Duncan 1976, Jordan et al. 2014, Molenberghs et al. 2016). Although there is debate as to whether group-based evaluations are moral (see Turiel 1983), group membership profoundly influences how adults evaluate others' morally relevant behaviors. Developmental research suggests that even infants and toddlers are sensitive to markers of group membership (e.g., language, race, culture-specific behaviors such as food preferences) (Liberman et al. 2017). Are early assessments of morally relevant behavior susceptible to the influence of group membership?

To address this question, Hamlin et al. (2013a) presented 9- and 14-month-olds with a protagonist who either shared or did not share their food preference. When the protagonist shared their food preference, infants preferred an agent who helped (versus hindered) the protagonist. When the protagonist did not share infants' food preference, infants instead preferred an agent who hindered the protagonist. Likewise, Burns & Sommerville (2014) presented White 15-month-olds with resource distributions involving people who either shared or did not share toddlers' racial backgrounds. When both distributors were White, toddlers preferred the fair over the unfair distributor. When the fair distributor was Asian and the unfair distributor was White, toddlers did not prefer the fair distributor. Moreover, when one recipient was Asian and the other White, toddlers more strongly preferred the fair distributor over the unfair distributor when unfair distribution meant that the Asian recipient would get more resources. Thus, ingroup preferences impact infants' and toddlers' evaluations of morally relevant behavior across both the help/hinder and fairness domains.

Other studies suggest that group membership similarly influences infants' and toddlers' expectations for morally relevant behavior. As reviewed above, when agents' groups are not specified, infants and toddlers have not demonstrated default expectations that agents will help versus hinder others. However, growing evidence suggests that when agents belong to groups, infants do hold expectations for helping versus hindering. For example, Pun et al. (2018) found that 6-to-12-month-olds expect agents who speak their native language (i.e., ingroup members) to help rather than hinder others but hold no such expectations of foreign language speakers (i.e., outgroup members). Likewise, Jin & Baillargeon (2017) found that 17-month-olds expect two agents who express that they share an identity (e.g., both identifying as "a bem") to help each other but hold no such expectations of agents who belong to different groups. Thus, infants and toddlers appear to expect helping behaviors among ingroup (but not outgroup) members.

In addition to ingroup helping, infants and toddlers appear to expect ingroup favoritism rather than fairness. As extensively reviewed above, when groups are not specified, infants and toddlers expect agents to distribute resources equally. When a distributor and one, but not a second, recipient belong to the same group, however, 1.5- and 2.5-year-olds expect a distributor to favor the ingroup recipient (Bian et al. 2018). In sum, group membership appears to influence infants' and toddlers' evaluations and expectations of helping, hindering, and fairness. Although this susceptibility to group membership may suggest immaturity in these assessments of morally relevant

behavior, it is consistent with adults' own morally relevant behaviors and cognitions across the world (Curry et al. 2019).

HOW EARLY UNDERSTANDING CONNECTS TO PROSOCIAL BEHAVIOR AND GIVES RISE TO LATER MORAL REASONING

What are the broader implications of infants' and toddlers' responses to morally relevant behaviors and intentions? Do these responses have any meaningful implications for children's prosocial behaviors and their later moral development? Observing such continuities would provide support for the possibility that these early-emerging tendencies play a meaningful role in moral development more broadly. In this section, we review evidence that infants' and toddlers' understanding and evaluations are related to their own prosocial behavior, that older children's moral responses align with infants' evaluations, and that early individual differences in social and moral responding predict subsequent social and moral functioning.

How Early Understanding Connects to Prosocial Behavior

A large body of research has found that in the second year, infants and toddlers will help, share, and comfort others (e.g., Dunfield et al. 2011, Warneken & Tomasello 2006). These prosocial behaviors may reflect a moral core: capacities not only for understanding and evaluation but also for action (Hamlin 2013b; see also Vaish & Tomasello 2014; cf. Dahl 2018). If so, then infants' and toddlers' understanding of others' morally relevant behaviors should be related to their own prosocial behaviors.

Evidence in support of this possibility has come from research examining infants' and toddlers' expectations of morally relevant behavior. In one study, individual differences in 15-month-olds' expectations that others will distribute resources fairly versus unfairly predicted whether the same toddlers would share an attractive toy with others (Schmidt & Sommerville 2011). In another study, infants who demonstrated stronger expectations of fairness also (according to caregivers' reports) shared more with others (Ziv & Sommerville 2017). These findings link infants' and toddlers' understanding of fairness to their own distributive behaviors.

Recent research has begun examining how an understanding of others' helping actions relates to one's own helping behavior. As reviewed above, infants and toddlers do not appear to have default expectations that agents will help versus hinder others. However, when two potential recipients differ in need (e.g., because only one can access a desired toy on its own), infants and toddlers appear to expect agents to help the recipient in more need (Köster et al. 2016). Building on these findings, Köster et al. (2019) found that among 16-month-olds with sufficient motor abilities, those who more strongly expected agents to help those in greater need were more likely to help others themselves. These findings link toddlers' understanding of helping to their own helpful behavior.

In sum, infants' and toddlers' sensitivity to morally relevant behavior appears connected to the same children's prosocial behavior. These findings suggest a unified moral core. How might this moral core relate to later social and moral development?

How Early Understanding Compares to Preschoolers' Evaluations: Cross-Sectional Evidence

To explore whether preschoolers' explicit judgments of helpers and hinderers align with infants' evaluations, Van de Vondervoort & Hamlin (2017) presented 3-to-5-year-old children with the helper/hinderer displays from Hamlin & Wynn's (2011) studies. They found that children preferred helpers, reported that helpers were nicer, and said that hinderers should be punished. Thus,

children's verbal responses were consistent with infants' evaluations, supporting the idea that the moral core serves as a foundation for later moral development.

Whereas infants' evaluations appear consistent with children's verbal evaluations of simple helper/hinderer displays, the evidence that infants and toddlers engage in intention-based evaluations stands in contrast with evidence that young children find it equally bad to cause negative outcomes accidentally versus intentionally (Margoni & Surian 2016, Piaget 1932). Although infants' and toddlers' intention-based evaluations may appear to conflict with young children's outcome-based evaluations, there have now been at least three accounts of this discrepancy. First, tasks designed for young children may be more cognitively demanding than those designed for infants (Margoni & Surian 2016, 2020). Second, tasks that are designed for young children may have a confound: Agents who engage in accidental behaviors may do so negligently or carelessly (i.e., they could or should have known better) and may therefore be held morally blameworthy for any negative consequences (Nobes et al. 2009; see also Woo et al. 2017). Third, whereas verbal tasks for young children focus on action acceptability, naughtiness, and punishment, nonverbal tasks for infants examine whom participants would rather engage with (e.g., via reaching or looking). Recent studies suggest that young children are more sensitive to others' intentions when tasks are framed around social engagement (e.g., whom to invite to a party, who is nicer) (Martin et al. 2021; Van de Vondervoort & Hamlin 2018). Taken together, these accounts suggest that there is continuity in intention-based evaluations between infancy and early childhood. Infants' sensitivity to prosocial and antisocial intentions may thus give rise to intention-based evaluations in early childhood.

How Early Understanding Relates to Later Social and Moral Functioning: Longitudinal Evidence

While research has revealed continuity between early nonverbal evaluations and later verbal evaluations, most of this research has been cross sectional, wherein different groups of children are tested in infancy versus later childhood. Is there continuity between evaluations of morally relevant behavior during infancy and later social and moral functioning within individual children?

To begin addressing these questions, Tan et al. (2018) examined the extent to which children's responses in studies probing their evaluations and expectations of helping and hindering as infants predict social and moral functioning in the same children during preschool. Tan et al. found that evaluations and expectations in infancy, but not habituation rates or fussiness, uniquely predicted moral and social functioning in preschool. Of particular interest, they assessed caregivers' ratings of preschoolers' callous-unemotional traits (e.g., whether children care about others and have a sense of right and wrong); callous-unemotional traits have been linked to children's own antisocial behaviors and conceptualization of morality (Feilhauer et al. 2013, Frick & White 2008). Here, children who were more likely to prefer more prosocial agents as infants were less likely to show callous-unemotional traits at 4 years; indeed, infants' responding predicted callous-unemotional traits but not other measures of more general social functioning. These findings suggest that infants' responses in studies presenting them with morally relevant behavior may have long-term implications for development. Future research probing whether individual differences in social and moral responding persist throughout infancy, childhood, and adulthood will further address the possibility that mature moral reasoning and behavior is rooted in a moral core.

SUMMARY AND OPEN QUESTIONS

A large body of evidence suggests that infants and toddlers understand and evaluate agents who engage in morally relevant behaviors. Growing evidence suggests that this understanding and

these evaluations of prosocial and antisocial agents are social and based on an understanding of the mental states of those receiving and those performing prosocial and antisocial behaviors. This evidence is consistent with the possibility that there is a moral core upon which socialization and active learning build to support mature social and moral reasoning later in life.

Key questions remain concerning the breadth and basis of infants' and toddlers' understanding of morally relevant behavior. First, although the evidence that we have reviewed is consistent with the possibility that there is a moral core, how do infants' and toddlers' environments support their understanding of prosocial and antisocial agents? For instance, Benton & Lapan (2022) have proposed that from observing others interact, infants learn about what behaviors are likely to lead to further opportunities for social interaction and suggest that infants' preferences for prosocial over antisocial agents are based on expectations of who is likely to interact with infants themselves. Regardless of whether learning leads infants to understand morally relevant behavior, better elucidating the social and moral environments of infants is a key step for future work. How often are young infants themselves the targets of prosocial versus antisocial behaviors? How often do they observe such behaviors between third parties? When they do observe these actions, who is performing them? Indeed, to the extent that infants' expectations and preferences stem from such experiences, variation in these experiences should predict their understanding and evaluations.

Second, although behavioral measures have produced largely consistent evidence across early childhood, and although early expectations and evaluations have been linked to prosocial behavior and to later development, looking and reaching—measures used with infants and toddlers—are ambiguous responses that may reflect a variety of different underlying mechanisms. For example, does preferential looking and reaching reflect social evaluation or other processes (e.g., curiosity)? Research examining young children's brain activity has begun to shed light on the neural mechanisms underlying infants' and toddlers' understanding and evaluations of morally relevant behavior. For instance, electroencephalogram research has found that helping and hindering scenarios trigger neural responses linked to approach and avoidance motivations in infants and toddlers and that viewing agents who have helped versus hindered elicits differential activation in neural mechanisms involved in social perception (Cowell & Decety 2015, Gredebäck et al. 2015, Tan & Hamlin 2022a). We look forward to research exploring children's brain activity in relation to other kinds of morally relevant behavior, which will help to reveal whether infants' understanding and evaluations of helpers and hinderers, fair and unfair distributors, and protectors and aggressors depend on the same or different neural mechanisms.

CONCLUSION

How does morality develop? Here, we have made the case that infants are far from morally blank slates. Rather, there is evidence for a moral core: early-emerging capacities for making sense of morally relevant behaviors. These capacities appear to be rooted in an understanding of others' goals and intentions and are sensitive to group membership. This moral core likely evolved to support the selection of appropriate cooperative partners and, combined with socialization and active learning processes, may give rise to human adults' moral sense.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Ashley Thomas, the Cambridge Writing Group, and Michael Gajda for their helpful feedback on an early version of this review, and we thank Leher Singh for sharing descriptive data from Singh (2020). B.W. is funded by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada (752-2020-0474). E.T. and J.K.H. were funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-03775).

LITERATURE CITED

Alexander JM. 2007. *The Structural Evolution of Morality*. Cambridge, MA: Cambridge Univ. Press

Aslin RN. 2007. What's in a look? *Dev. Sci.* 10(1):48–53

Atari M, Haidt J, Graham J, Koleva S, Stevens ST, Dehghani M. 2022. Morality beyond the WEIRD: how the nomological network of morality varies across cultures. PsyArXiv. **https://doi.org/10.31234/osf.io/q6c9r**

Augustine ME, Stifter CA. 2015. Temperament, parenting, and moral development: specificity of behavior and context. *Soc. Dev.* 24(2):285–303

Axelrod R. 1984. *The Evolution of Cooperation*. New York: Basic Books

Bandura A. 2014. Social cognitive theory of moral thought and action. In *Handbook of Moral Behavior and Development*, Vol. 1: *Theory*, ed. WM Kurtines, JL Gerwitz, pp. 45–103. New York: Psychol. Press

Benton DT, Lapan C. 2022. Moral masters or moral apprentices? A connectionist account of sociomoral evaluation in preverbal infants. *Cogn. Dev.* 62:101164

Bian L, Sloane S, Baillargeon R. 2018. Infants expect ingroup support to override fairness when resources are limited. *PNAS* 115(11):2705–10

Brown DE. 1991. *Human Universals*. Philadelphia, PA: Temple Univ. Press

Burns MP, Sommerville J. 2014. "I pick you": the impact of fairness and race on infants' selection of social partners. *Front. Psychol.* 5:93

Buon M, Jacob P, Margules S, Brunet I, Dutat M, et al. 2014. Friend or foe? Early social evaluation of human interactions. *PLOS ONE* 9(2):e88612

Buyukozer Dawkins M, Sloane S, Baillargeon R. 2019. Do infants in the first year of life expect equal resource allocations? *Front. Psychol.* 10:116

Carey S. 2009. *The Origin of Concepts*. Oxford, UK: Oxford Univ. Press

Chae JJK, Song H. 2018. Negativity bias in infants' expectations about agents' dispositions. *Br. J. Dev. Psychol.* 36(4):620–33

Colombo J, Mitchell DW. 2009. Infant visual habituation. *Neurobiol. Learn. Mem.* 92(2):225–34

Cosmides L, Tooby J. 1992. Cognitive adaptations for social exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. JH Barkow, L Cosmides, J Tooby, pp. 163–228. New York: Oxford Univ. Press

Cowell JM, Decety J. 2015. Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *PNAS* 112(41):12657–62

Curry OS, Mullins DA, Whitehouse H. 2019. Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Curr. Anthropol.* 60(1):47–69

Cushman F, Young L, Hauser M. 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychol. Sci.* 17(12):1082–89

Dahl A. 2016. Mothers' insistence when prohibiting infants from harming others in everyday interactions. *Front. Psychol.* 7:1448

Dahl A. 2018. New beginnings: an interactionist and constructivist approach to early moral development. *Hum. Dev.* 61(4–5):232–47

Dahl A, Campos JJ. 2013. Domain differences in early social interactions. *Child Dev.* 84(3):817–25

Darwin C. 1871. *The Descent of Man, and Selection in Relation to Sex*. New York: Appleton & Co.

DesChamps TD, Eason AE, Sommerville JA. 2016. Infants associate praise and admonishment with fair and unfair individuals. *Infancy* 21(4):478–504

Duncan BL. 1976. Differential social perception and attribution of intergroup violence: testing the lower limits of stereotyping of Blacks. *J. Pers. Soc. Psychol.* 34(4):590–98

Dunfield K, Kuhlmeier VA, O'Connell L, Kelley E. 2011. Examining the diversity of prosocial behavior: helping, sharing, and comforting in infancy. *Infancy* 16(3):227–47

Dunn J. 2014. Moral development in early childhood and social interaction in the family. See Killen & Smetana 2014, pp. 135–59

Enright EA, Gweon H, Sommerville JA. 2017. 'To the victor go the spoils': infants expect resources to align with dominance structures. *Cognition* 164:8–21

Fawcett C, Liszkowski U. 2012. Infants anticipate others' social preferences. *Infant Child Dev.* 21(3):239–49

Fehr E, Fischbacher U. 2003. The nature of human altruism. *Nature* 425(6960):785–91

Fehr E, Gächter S. 2002. Altruistic punishment in humans. *Nature* 415(6868):137–40

Feilhauer J, Cima M, Benjamins C, Muris P. 2013. Knowing right from wrong, but just not always feeling it: relations among callous-unemotional traits, psychopathological symptoms, and cognitive and affective morality judgments in 8- to 12-year-old boys. *Child Psychiatry Hum. Dev.* 44(6):709–16

Frick PJ, White SF. 2008. Research review: the importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *J. Child Psychol. Psychiatry* 49(4):359–75

Geraci A. 2020a. How do toddlers evaluate defensive actions toward third parties? *Infancy* 25(6):910–26

Geraci A. 2020b. How evaluation of protective third-party interventions and the relationship context interact at 21 months. *Eur. J. Dev. Psychol.* 17(4):556–77

Geraci A, Simion F. 2021. Evaluation of prosocial actions performed by dynamic shapes at 17 months of age. *Eur. J. Dev. Psychol*. In press. **https://doi.org/10.1080/17405629.2021.1957823**

Geraci A, Simion F, Surian L. 2022. Infants' intention-based evaluations of distributive actions. *J. Exp. Child Psychol.* 220:105429

Geraci A, Surian L. 2011. The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Dev. Sci.* 14(5):1012–20

Gintis H, Henrich J, Bowles S, Boyd R, Fehr E. 2008. Strong reciprocity and the roots of human morality. *Soc. Justice Res.* 21(2):241–53

Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, Ditto PH. 2011. Mapping the moral domain. *J. Pers. Soc. Psychol.* 101(2):366–85

Gredebäck G, Kaduk K, Bakker M, Gottwald J, Ekberg T, et al. 2015. The neuropsychology of infants' prosocial preferences. *Dev. Cogn. Neurosci.* 12:106–13

Greene J, Haidt J. 2002. How (and where) does moral judgment work? *Trends Cogn. Sci.* 6(12):517–23

Grusec JE, Chaparro MP, Johnston M, Sherman A. 2014. The development of moral behavior from a socialization perspective. See Killen & Smetana 2014, pp. 113–34

Gweon H. 2021. Inferential social learning: cognitive foundations of human social learning and teaching. *Trends Cogn. Sci.* 25(10):896–910

Hamlin JK. 2013a. Failed attempts to help and harm: intention versus outcome in preverbal infants' social evaluations. *Cognition* 128(3):451–74

Hamlin JK. 2013b. Moral judgment and action in preverbal infants and toddlers: evidence for an innate moral core. *Curr. Dir. Psychol. Sci.* 22(3):186–93

Hamlin JK. 2015. The case for social evaluation in preverbal infants: gazing toward one's goal drives infants' preferences for Helpers over Hinderers in the hill paradigm. *Front. Psychol.* 5:1563

Hamlin JK, Mahajan N, Liberman Z, Wynn K. 2013a. Not like me = bad: Infants prefer those who harm dissimilar others. *Psychol. Sci.* 24(4):589–94

Hamlin JK, Ullman T, Tenenbaum J, Goodman N, Baker C. 2013b. The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Dev. Sci.* 16(2):209–26

Hamlin JK, Wynn K. 2011. Young infants prefer prosocial to antisocial others. *Cogn. Dev.* 26(1):30–39

Hamlin JK, Wynn K, Bloom P. 2007. Social evaluation by preverbal infants. *Nature* 450(7169):557–59

Hamlin JK, Wynn K, Bloom P. 2010. Three-month-olds show a negativity bias in their social evaluations. *Dev. Sci.* 13(6):923–29

Hamlink JK, Wynn K, Bloom P, Mahajan N. 2011. How infants and toddlers react to antisocial others. *PNAS* 108(50):19931–36

Hawkes K. 2014. Primate sociality to human cooperation. *Hum. Nat.* 25(1):28–48
Heider F. 1958. *The Psychology of Interpersonal Relations*. New York: Psychol. Press
Henrich N, Henrich JP. 2007. *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. New York: Oxford Univ. Press
Holvoet C, Arciszewski T, Scola C, Picard D. 2018. Infants' visual preferences for prosocial behavior and other-race characters at 6 months: an eye-tracking study. *SAGE Open* 8(2). **https://doi.org/10.1177/2158244018784993**
Hood BM, Willen JD, Driver J. 1998. Adult's eyes trigger shifts of visual attention in human infants. *Psychol. Sci.* 9(2):131–34
Hrdy SB. 1999. *Mother Nature: A History of Mothers, Infants, and Natural Selection*. New York: Pantheon
Hrdy SB. 2009. *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, MA: Harvard Univ. Press
Jin K, Baillargeon R. 2017. Infants possess an abstract expectation of ingroup support. *PNAS* 114(31):8199–204
Jordan JJ, McAuliffe K, Warneken F. 2014. Development of in-group favoritism in children's third-party punishment of selfishness. *PNAS* 111(35):12710–15
Joyce R. 2007. *The Evolution of Morality*. Cambridge, MA: MIT Press
Kanakogi Y, Inoue Y, Matsuda G, Butler D, Hiraki K, Myowa-Yamakoshi M. 2017. Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat. Hum. Behav.* 1(2):37
Kanakogi Y, Okumura Y, Inoue Y, Kitazaki M, Itakura S. 2013. Rudimentary sympathy in preverbal infants: preference for others in distress. *PLOS ONE* 8(6):e65292
Katz LD, ed. 2000. *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives*. Bowling Green, OH: Impr. Acad.
Killen M, Smetana G, eds. 2014. *Handbook of Moral Development*. New York: Psychol. Press. 2nd ed.
Kochanska G. 2002. Mutually responsive orientation between mothers and their young children: a context for the early development of conscience. *Curr. Dir. Psychol. Sci.* 11(6):191–95
Köster M, Itakura S, Omori M, Kärtner J. 2019. From understanding others' needs to prosocial action: motor and social abilities promote infants' helping. *Dev. Sci.* 22(6):e12804
Köster M, Ohmer X, Nguyen TD, Kärtner J. 2016. Infants understand others' needs. *Psychol. Sci.* 27(4):542–48
Kuczynski L, Knafo A. 2014. Innovation and continuity in socialization, internalization, and acculturation. See Killen & Smetana 2014, pp. 93–112
Kuhlmeier V, Wynn K, Bloom P. 2003. Attribution of dispositional states by 12-month-olds. *Psychol. Sci.* 14(5):402–8
Lee W, Kim EY, Song H. 2020. Do infants expect others to be helpful? *Br. J. Dev. Psychol.* 38(3):478–90
Lee Y, Yun JE, Kim EY, Song H. 2015. The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLOS ONE* 10(9):e0135588
Liberman Z, Woodward AL, Kinzler KD. 2017. The origins of social categorization. *Trends Cogn. Sci.* 21(7):556–68
Loheide-Niesmann L, de Lijster J, Hall R, van Bakel H, Cima M. 2021. Toddlers' preference for prosocial versus antisocial agents: no associations with empathy or attachment security. *Soc. Dev.* 30(2):410–27
Lucca K, Pospisil J, Sommerville JA. 2018. Fairness informs social decision making in infancy. *PLOS ONE* 13(2):e0192848
Malle BF. 1999. How people explain behavior: a new theoretical framework. *Pers. Soc. Psychol. Rev.* 3(1):23–48
Margoni F, Surian L. 2016. Explaining the U-shaped development of intent-based moral judgments. *Front. Psychol.* 7:219
Margoni F, Surian L. 2018. Infants' evaluation of prosocial and antisocial agents: a meta-analysis. *Dev. Psychol.* 54(8):1445–55
Margoni F, Surian L. 2020. Conceptual continuity in the development of intent-based moral judgment. *J. Exp. Child Psychol.* 194:104812
Martin JW, Leddy K, Young L, McAuliffe K. 2021. An earlier role for intent in children's partner choice versus punishment. *J. Exp. Psychol. Gen.* 151(3):597–612
McNamara RA, Willard AK, Norenzayan A, Henrich J. 2019. Weighing outcome vs. intent across societies: how cultural models of mind shape moral reasoning. *Cognition* 182:95–108

Meristo M, Strid K, Surian L. 2016. Preverbal infants' ability to encode the outcome of distributive actions. *Infancy* 21(3):353–72

Molenberghs P, Gapp J, Wang B, Louis WR, Decety J. 2016. Increased moral sensitivity for outgroup perpetrators harming ingroup members. *Cereb. Cortex* 26(1):225–33

Nobes G, Panagiotaki G, Pawson C. 2009. The influence of negligence, intention, and outcome on children's moral judgments. *J. Exp. Child. Psychol.* 104(4):382–97

Piaget J. 1932. *The Moral Judgment of the Child.* London: Routledge & K. Paul

Powell LJ. 2019. Imitation: neither instinct nor gadget, but a cultural starting point? *Behav. Brain Sci.* 42:e180

Powell LJ, Spelke ES. 2018. Third-party preferences for imitators in preverbal infants. *Open Mind* 2(2):61–71

Pun A, Ferera M, Diesendruck G, Hamlin JK, Baron AS. 2018. Foundations of infants' social group evaluations. *Dev. Sci.* 21(3):e12586

Rottman J, Young L. 2015. Mechanisms of moral development. In *The Moral Brain: A Multidisciplinary Perspective*, ed. J Decety, T Wheatley, pp. 123–42. Cambridge, MA: MIT Press

Salvadori E, Blazsekova T, Volein A, Karap Z, Tatone D, et al. 2015. Probing the strength of infants' preference for helpers over hinderers: two replication attempts of Hamlin and Wynn 2011. *PLOS ONE* 10(11):e0140570

Saxe R. 2016. Moral status of accidents. *PNAS* 113(17):4555–57

Scarf D, Imuta K, Colombo M, Hayne H. 2012a. Golden Rule or valence matching? Methodological problems in Hamlin et al. *PNAS* 109(22):E1426

Scarf D, Imuta K, Colombo M, Hayne H. 2012b. Social evaluation or simple association? Simple associations may explain moral reasoning in infants. *PLOS ONE* 7(8):e42698

Schlingloff L, Csibra G, Tatone D. 2020. Do 15-month-old infants prefer helpers? A replication of Hamlin et al. 2007. *R. Soc. Open Sci.* 7(4):191795

Schmidt MFH, Sommerville JA. 2011. Fairness expectations and altruistic sharing in 15-month-old human infants. *PLOS ONE* 6(10):e23223

Scola C, Holvoet C, Arciszewski T, Picard D. 2015. Further evidence for infants' preference for prosocial over antisocial behaviors. *Infancy* 20(6):684–92

Singh L. 2020. Bilingual infants are more sensitive to morally relevant social behavior than monolingual infants. *J. Cogn. Dev.* 21(5):631–50

Sloane S, Baillargeon R, Premack D. 2012. Do infants have a sense of fairness? *Psychol. Sci.* 23(2):196–204

Smetana JG. 1989. Adolescents' and parents' reasoning about actual family conflict. *Child Dev.* 60(5):1052–67

Smetana JG. 2006. Social-cognitive domain theory: consistencies and variations in children's moral and social judgments. In *Handbook of Moral Development*, ed. M Killen, G Smetana, pp. 119–53. Mahwah, NJ: Lawrence Erlbaum Assoc. 1st ed.

Smetana JG, Jambon M, Ball C. 2014. The social domain approach to children's moral and social judgments. See Killen & Smetana 2014, pp. 23–45

Smetana JG, Rote WM, Jambon M, Tasopoulos-Chan M, Villalobos M, Comer J. 2012. Developmental changes and individual differences in young children's moral judgments. *Child Dev.* 83(2):683–96

Sommerville JA, Schmidt MFH, Yun J, Burns M. 2013. The development of fairness expectations and prosocial behavior in the second year of life. *Infancy* 18(1):40–66

Sommerville JA, Woodward AL. 2005. Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition* 95(1):1–30

Spelke ES, Kinzler KD. 2007. Core knowledge. *Dev. Sci.* 10(1):89–96

Sterelny K, Joyce R, Calcott B, Fraser B, eds. 2013. *Cooperation and Its Evolution*. Cambridge, MA: MIT Press

Tafreshi D, Thompson JJ, Racine TP. 2014. An analysis of the conceptual foundations of the infant preferential looking paradigm. *Hum. Dev.* 57(4):222–40

Tan E, Hamlin JK. 2022a. Infants' neural responses to helping and hindering scenarios. *Dev. Cogn. Neurosci.* 54:101095

Tan E, Hamlin JK. 2022b. Mechanisms of social evaluation in infancy: a preregistered exploration of infants' eye-movement and pupillary responses to prosocial and antisocial events. *Infancy* 27(2):255–76

Tan E, Mikami AY, Hamlin JK. 2018. Do infant sociomoral evaluation and action studies predict preschool social and behavioral adjustment? *J. Exp. Child Psychol.* 176:39–54

Ting F, Dawkins MB, Stavans M, Baillargeon R. 2019. Principles and concepts in early moral cognition. In *The Social Brain: A Developmental Perspective*, ed. J Decety, pp. 41–65. Cambridge, MA: MIT Press

Tomasello M, Carpenter M. 2007. Shared intentionality. *Dev. Sci.* 10(1):121–25

Trivers RL. 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* 46(1):35–57

Turiel E. 1983. *The Development of Social Knowledge: Morality and Convention*. Cambridge, UK: Cambridge Univ. Press

Turiel E. 2014. Epistemology, development, and social opposition. See Killen & Smetana 2014, pp. 3–22

Vaish A, Grossmann T, Woodward A. 2008. Not all emotions are created equal: the negativity bias in social-emotional development. *Psychol. Bull.* 134(3):383–403

Vaish A, Tomasello M. 2014. The early ontogeny of human cooperation and morality. See Killen & Smetana 2014, pp. 279–98

Van de Vondervoort JW, Aknin LB, Kushnir T, Slevinsky J, Hamlin JK. 2018. Selectivity in toddlers' behavioral and emotional reactions to prosocial and antisocial others. *Dev. Psychol.* 54(1):1–14

Van de Vondervoort JW, Hamlin JK. 2017. Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *J. Exp. Child Psychol.* 164:136–51

Van de Vondervoort JW, Hamlin JK. 2018. Preschoolers focus on others' intentions when forming sociomoral judgments. *Front. Psychol.* 9:1851

Vaporova E, Zmyj N. 2020. Social evaluation and imitation of prosocial and antisocial agents in infants, children, and adults. *PLOS ONE* 15(9):e0235595

Warneken F, Tomasello M. 2006. Altruistic helping in human infants and young chimpanzees. *Science* 311(5765):1301–3

Woo BM, Spelke ES. 2020. How to help best: infants' changing understanding of multistep actions informs their evaluations of helping. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, ed. S Denison, M Mack, Y Xu, BC Armstrong, pp. 384–90. Austin, TX: Cogn. Sci. Soc.

Woo BM, Spelke ES. 2022a. Eight-month-old infants' social evaluations of agents who act on false beliefs. Paper presented at the 44th Annual Conference of the Cognitive Science Society, Toronto, July 28

Woo BM, Spelke ES. 2022b. Infants and toddlers leverage their understanding of action goals to evaluate agents who help others. PsyArXiv. **https://doi.org/10.31234/osf.io/mtprn**

Woo BM, Spelke ES. 2022c. Toddlers' social evaluations of agents who act on false beliefs. *Dev. Sci.* In press.

Woo BM, Steckler CM, Le DT, Hamlin JK. 2017. Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition* 168:154–63

Yoo HN, Smetana JG. 2021. Associations among child temperament, parenting, and young children's moral and conventional understanding: the moderating role of self-regulation. *Soc. Dev.* In press. **https://doi.org/10.1111/sode.12571**

Young L, Cushman F, Hauser M, Saxe R. 2007. The neural basis of the interaction between theory of mind and moral judgment. *PNAS* 104(20):8235–40

Ziv T, Sommerville JA. 2017. Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Dev.* 88(6):1930–51

Ziv T, Whiteman JD, Sommerville JA. 2021. Toddlers' interventions toward fair and unfair individuals. *Cognition* 214:104781