Principles and Concepts in Early Moral Cognition

Fransisca Ting[1], Melody Buyukozer Dawkins[1], Maayan Stavans[2], Renée Baillargeon[1]

[1]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

[2]Department of Cognitive Science, Cognitive Development Center, Central European University, 1051 Budapest, Hungary

Corresponding author:

Fransisca Ting, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820; phone: (217) 333-5988; fax: (217) 244-5876; email: fting2@illinois.edu.

Acknowledgment:

Preparation of this chapter was supported by a grant from the John Templeton Foundation (to R.B.).

Citation:

Ting, F., Buyukozwer Dawkins, M., Stavans, M., & Baillargeon, R. (in press). Principles and concepts in early moral cognition. To appear in J. Decety (Ed.), *The social brain: A developmental perspective*. Cambridge, MA: MIT Press.

Abstract

According to Graham and colleagues (2013), a “first draft” of moral cognition emerges early and universally in development and is then gradually revised by experience and culture. In this chapter, we explore some of the moral principles and concepts that might be included in this initial draft of morality. First, we review evidence that at least four moral principles guide infants’ reasoning about how individuals should act toward others: fairness, harm avoidance, ingroup support, and authority. These principles regulate actions at different levels of the social landscape and interact in various ways. Next, we report recent results concerning four moral concepts that support infants’ principle-based expectations: moral obligation, moral status, moral circle, and moral character. Together, these findings indicate that early human moral cognition is remarkably sophisticated and provides a rich foundation for infants’ adaptation to their social worlds.

1. Introduction

Over the past two decades, a wealth of research has examined infants' evaluations of social actions, to uncover both how these evaluations are formed and how they affect infants' responses to others. Initially, it appeared as though early sociomoral evaluations reflected mainly infants' ability to distinguish between positive and negative actions (Bloom & Wynn, 2016; Hamlin, 2013b). In a series of experiments, for example, Hamlin and colleagues familiarized infants ages 3–21 months to scenarios depicting interactions among non-human individuals (e.g., blocks with eyes, non-verbal puppets; Hamlin, 2013a, 2014, 2015; Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007, 2010; Hamlin, Wynn, Bloom, & Mahajan, 2011). Each scenario involved a positive event, in which one character acted positively toward a protagonist (e.g., helped it reach the top of a steep hill, returned a ball it had dropped), and a negative event, in which another character acted negatively toward the same protagonist (e.g., knocked it down to the bottom of the hill, stole its ball). Across ages and scenarios, infants tended to look equally at the positive and negative events, suggesting that they did not expect the characters either to help the protagonist achieve its goal or to refrain from harming it. These negative results did not stem from infants' inability to understand the events presented: Following the familiarization trials, 3–11-month-olds tested with a social-preference task chose the positive over the negative character (Hamlin, 2015; Hamlin et al., 2007, 2010; Hamlin & Wynn, 2011; for a meta-analysis, see Margoni & Surian, 2018); 8-month-olds did so even if the characters' positive and negative actions were unsuccessful, indicating that infants attended primarily to the characters' intentions in evaluating their actions (Hamlin, 2013a); 10-month-olds tested with a violation-of-expectation task expected the protagonist to also prefer the positive character and detected a violation (as indexed by longer looking times) when the protagonist approached the negative character instead (Hamlin et al., 2007); and 21-month-olds chose the positive character when asked to give away a treat, but chose the negative character when asked to take away a treat (Hamlin et al., 2011). Together, these results suggested that early sociomoral sensitivities included no particular expectations about how individuals would act toward others. Nevertheless, infants possessed notions of welfare and harm that enabled them to evaluate actions as positive (beneficial to the

targets of the actions), neutral, or negative (detrimental to the targets of the actions). These evaluations, in turn, drove affiliative attitudes: Infants preferred, and expected others to prefer, individuals who produced positive actions over individuals who produced negative actions.

In time, however, it became clear that infants possess richer moral sensitivities than was initially thought. One of the turning points in this research came from new evidence that (a) infants do hold expectations about how individuals will act toward others, but (b) these expectations are often context-sensitive and arise only when specific pre-conditions are met. For example, 13–17-month-olds viewed providing help as obligatory when individuals belonged to the same social group, but as optional when individuals belonged to different groups or when group memberships were unspecified (Jin & Baillargeon, 2017; Ting, He, & Baillargeon, 2019a). Similarly, 17-month-olds viewed directly intervening in a transgression against an ingroup victim as obligatory for a group leader, but as optional for a non-leader equal in rank to the victim (Stavans & Baillargeon, in press).

These and other findings (reviewed in the next sections) have led us to propose a new characterization of infant moral cognition that rests on five assumptions. First, as is the case for early physical and psychological reasoning (Baillargeon, Li, Gertner, & Wu, 2011; Baillargeon, Scott, & Bian, 2016), early sociomoral reasoning is guided by a skeletal framework of principles and concepts that emerges early and universally in development (Baillargeon et al., 2015; Buyukozer Dawkins, Ting, Stavans, & Baillargeon, in press). Second, early sociomoral principles include *fairness, harm avoidance, ingroup support*, and *authority*—all of which have long been the focus, explicitly or implicitly, of research across the social sciences (Baumard, André, & Sperber, 2013; Brewer, 1999; Dupoux & Jacob, 2007; Graham et al., 2013; Rai & Fiske, 2011; Shweder, Much, Mahapatra, & Park, 1997). Each principle is normative in nature and specifies what is obligatory and forbidden in its context of application. Third, each principle is tied to a particular social distinction or structure. Thus, fairness, the broadest of the principles, applies to interactions among any individuals (i.e., entities with moral status); harm avoidance applies to interactions among individuals who are identified as members of the same moral circle (e.g., humans); ingroup support applies to

interactions among individuals in a moral circle who are identified as members of the same social group (e.g., sports team); and authority applies to interactions among individuals in a social group who are identified as leaders and followers (e.g., coach and players). Thus, with each successive social differentiation, additional expectations come into play that help regulate interactions at this new level. Fourth, when two or more principles apply to the same situation, they interact in various ways. In some cases, for example, one principle may intensify expectations set by another principle (e.g., even less harm is tolerated when individuals belong not only to the same moral circle but also to the same group; Ting, He & Baillargeon, 2019b); in other cases, two principles may suggest different courses of action and must then be rank-ordered (e.g., fairness may dictate that a resource be divided equally between ingroup and outgroup recipients, but ingroup support may dictate that it be reserved for ingroup recipients, particularly when it is scarce or otherwise valuable; Bian, Sloane & Baillargeon, 2018). Finally, the last assumption is that different cultures implement, stress, and rank-order the principles differently, resulting in the diverse moral landscape that exists in the world today.

Graham et al. (2013) described their work on moral foundations as "a theory about the universal first draft of the moral mind and about how that draft gets revised in variable ways across cultures" (p. 65). From this perspective, efforts to specify the principles and concepts that shape early sociomoral reasoning thus help shed light on "the universal first draft" of human moral cognition. In this chapter, we first review evidence that principles of fairness, harm avoidance, ingroup support, and authority guide early sociomoral expectations. Next, we discuss four key concepts implicated in these expectations: moral obligation, moral status, moral circle, and moral character.

2. **Moral Principles**

Fairness. When watching interactions among individuals, or entities with moral status, infants bring to bear an equity-based principle of fairness: All other things being equal, individuals are expected to give others their just deserts—that is, to treat them as they deserve to be treated in the situation at hand. Thus, in situations involving the distribution of *windfall resources*, 4–15-month-olds expected a distributor to divide items equally between two similar potential recipients (Buyukozer Dawkins, Sloane, & Baillargeon, 2019; Meristo, Strid, & Surian, 2016; Schmidt & Sommerville, 2011); 10–

15-month-olds found it unexpected when an unfair (but not a fair) distributor was rewarded or praised (DesChamps, Eason, & Sommerville, 2015; Meristo & Surian, 2013); 13–17-month-olds preferred a fair over an unfair distributor when asked to choose between them (Burns & Sommerville, 2014; Geraci & Surian, 2011; Lucca, Pospisil, & Sommerville, 2018); 20–30-month-olds chose to help a fair as opposed to an unfair distributor (Surian & Franchin, 2017); and 21-month-olds took into account recipients' pre-existing resources and expected resource-poor recipients to receive a larger share than resource-rich recipients (Buyukozer Dawkins & Baillargeon, 2019). In situations involving the dispensation of *rewards* for efforts, 10-month-olds found it unexpected when an experimenter gave equal praise to an assistant who had done a puzzle and an assistant who had not (Buyukozer Dawkins, Sloane, & Baillargeon, 2017); 17-month-olds expected a resource acquired by two workers to be shared according to the amount of effort each had exerted (Wang & Henderson, 2018); and 21-month-olds found it unexpected when an experimenter gave the same reward to a worker who had done an assigned chore and a slacker who had done no work (Sloane, Baillargeon, & Premack, 2012). Finally, in a situation involving the meting out of *punishments*, 21-month-olds found it unexpected when two assistants were both punished even though only one of them had disobeyed an instruction (Buyukozer Dawkins et al., 2017). Together, these results provide converging evidence that from a young age, infants possess an equity-based expectation of fairness: Individuals are expected to receive the treatment they deserve in each situation, be it an equitable share of windfall resources, a reward commensurate with their efforts, or a punishment befitting their misdeeds.

Harm avoidance. When individuals belong to the same moral circle, infants bring to bear an abstract principle of harm avoidance that sets broad limits on the amount of unprovoked harm the individuals can inflict on one another. Thus, when watching interactions between two individuals, A and B, who belonged to the same moral circle but gave no indication of belonging to the same social group (e.g., two humans who belonged to different groups, two humans whose group memberships were unspecified, or two puppets with the power of speech—henceforth verbal

puppets—who belonged to different animal kinds[1]), 13–33-month-olds did not find it unexpected when A directed a relatively *mild* negative action toward B: for example, when A ignored B's need for instrumental assistance (Jin & Baillargeon, 2017), threw an object B needed on the floor (Ting et al., 2019a), crumpled a drawing done by B (Ting & Baillargeon, 2018a), or knocked down one block from one of three towers built by B (Ting et al., 2019b). However, infants did detect a violation when A's negative actions were *more intense*: for example, when A directed several mild negative actions toward B (Ting & Baillargeon, 2018a) or knocked down one of three towers built by B (Ting et al., 2019b). These last results contrast sharply with those obtained when A and B gave no cue of belonging to the same moral circle (e.g., two different non-verbal puppets or geometric figures), as the amount of unprovoked harm deemed acceptable was then much greater. Thus, 6–15-month-olds detected no violation when A repeatedly hit B (Kanakogi et al., 2017; Premack & Premack, 1997), growled at and fought with B over the possession of a toy (Rhodes, Hetherington, Brink & Wellman, 2015), or knocked B down a steep hill, causing it to roll end-over-end to the bottom of the hill (Hamlin, 2015). Together, these results suggest that from a young age, infants expect harm avoidance to apply within but not between moral circles.

Ingroup support. When individuals in a moral circle belong to the same social group, infants bring to bear an abstract principle of ingroup support. This principle carries numerous expectations that can be roughly divided into two sets related to ingroup care and loyalty. With respect to *ingroup care*, 4–12-month-olds expected a woman alone with a crying baby (who presumably belonged to the same group as the woman) to attempt to comfort the baby, and they found it unexpected when she ignored the baby instead (Jin, Houston, Baillargeon, Groh, & Roisman, 2018); and 13–29-month-olds expected a woman to provide help to another woman in need of instrumental assistance when the two belonged to the same group, but they held no

[1] Infants appear to view puppets with the power of speech as members of the human moral circle or "honorary humans" (e.g., Big Bird, Elmo). Moreover, just as infants assign humans to different social groups, based on various cues (e.g., Burns & Sommerville, 2014; Jin & Baillargeon, 2017; Liberman, Woodward, Sullivan, & Kinzler, 2016; Ting et al., 2019a), they assign verbal puppets from different animal kinds (e.g., English-speaking monkeys, giraffes, rabbits, and dogs) to different social groups (Bian et al., 2018; Ting & Baillargeon, 2018a).

expectation about the provision of help when the two women belonged to different groups or when their group memberships were unspecified (Jin & Baillargeon, 2017; Ting et al., 2019a). In situations invoking both harm avoidance and ingroup support, 13–33-month-olds expected individuals to refrain from *any* unprovoked harm toward ingroup members, and they therefore detected a violation when mild negative actions that were deemed permissible against outgroup members (e.g., failing to provide help, throwing a needed object on the floor, crumpling a drawing, or knocking down part of a tower) were directed at ingroup members (Jin & Baillargeon, 2017; Ting & Baillargeon, 2018a; Ting et al., 2019a, 2019b). In the same vein, 3-year-olds who heard stories about two novel social groups predicted that a harmful action (e.g., stealing a block) would be more likely to be directed at an outgroup as opposed to an ingroup victim (Rhodes, 2012). Finally, in studies of second-party punishment, 18-month-olds expected less retaliatory harm to be directed at an ingroup as opposed to an outgroup wrongdoer for the same transgression (Ting et al., 2019b); and in studies of indirect third-party punishment, 13–29-month-olds expected an individual who had observed a transgression against an ingroup victim to later refrain from providing instrumental assistance to the wrongdoer, to signal that transgressions against the group had adverse consequences and thus to help deter future transgressions (Ting et al., 2019a).

Turning to *ingroup loyalty*, 6–10-month-olds preferred a native speaker of their language over a foreign speaker (Kinzler, Dupoux, & Spelke, 2007), 12-month-olds preferred toys or snacks endorsed by a native speaker over those endorsed by a foreign speaker (Kinzler, Dupoux, & Spelke, 2012; Shutts, Kinzler, McKee, & Spelke, 2009), and 14–19-month-olds were more likely to imitate a novel conventional action modeled by a native as opposed to a foreign speaker (Buttelmann, Zmyj, Daum, & Carpenter, 2013; Howard, Henderson, Carrazza, & Woodward, 2015). Similarly, after watching two groups of non-human characters perform distinct novel conventional actions, 7–12-month-olds detected a violation if a member of one group chose to imitate the other group's conventional action (Powell & Spelke, 2013); 12-month-olds expected an individual to choose an ingroup over an outgroup member as a play partner (Bian & Baillargeon, 2019); and after watching adult characters soothe baby characters (Spokes & Spelke, 2017), 16-month-olds detected a violation if one baby chose as a play partner

a baby who had been soothed by a different adult (and hence presumably belonged to a different group) over a baby who had been soothed by the same adult (and hence presumably belonged to the same group). Finally, in resource-distribution situations invoking both fairness and ingroup support, 19–28-month-olds expected fairness to prevail when there were sufficient resources for all ingroup and outgroup individuals present, but they expected ingroup support to trump fairness when there were only enough resources for the ingroup individuals (Bian et al., 2018).

Together, these results provide converging evidence that from a young age, an abstract principle of ingroup support guides infants' reasoning about interactions within social groups and carries rich expectations of ingroup care and loyalty.

Authority. According to Rai and Fiske (2011), the principle of authority carries moral obligations for both leaders and followers: Leaders are expected "to lead, guide, direct, and protect" their followers, whereas followers are expected "to respect, obey, and pay deference" to their leaders (p. 63). Building on prior evidence that young infants detect power asymmetries (Pun, Birch, & Baron, 2016; Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011) and expect them to be stable over time and to extend across situations (Gazes, Hampton, & Lourenco, 2017; Mascaro & Csibra, 2012), recent research has begun to examine whether infants share some of the same authority-based expectations as adults. In experiments focusing on expectations about *leaders* (Stavans & Baillargeon, in press), 17-month-olds watched a group of three bear puppets who served as the protagonist, wrongdoer, and victim. The protagonist brought in two toys for the other bears to share, but the wrongdoer unfairly seized both toys, leaving none for the victim. The protagonist then either took one toy away from the wrongdoer and gave it to the victim (*intervention* event) or approached each bear in turn without re-distributing a toy (*non-intervention* event). Infants expected an intervention when the protagonist was portrayed as a leader, but they held no particular expectation for intervention when the protagonist was portrayed as a non-leader equal in rank to the other bears. In other experiments focusing on expectations about *followers* (Margoni, Baillargeon, & Surian, 2018), 21-month-olds saw computer-animated events in which three characters were playing in a field next to a house when a protagonist arrived and ordered them to go to bed; the characters then filed into the house and could be seen

through its front window. Next, the protagonist left the scene, and the characters either returned to the field (*disobedience* event) or remained in the house and went to sleep (*obedience* event). When the protagonist was portrayed as a leader, infants expected the characters to continue to obey her order after she left the scene; when the protagonist was portrayed as a bully, however, infants held no expectation about whether the characters would continue to obey her after she left. Together, these results suggest that by the second year of life, an abstract principle of authority guides infants' expectations about interactions between leaders and followers within social groups.

3. Moral Concepts

Embedded in our descriptions of early principle-based expectations are several key moral concepts. In this section, we discuss four such concepts: We consider whether infants understand *moral obligations*, on what entities they confer *moral status*, how they identify *moral circles*, and how they evaluate individuals' *moral characters*.

Moral obligations. We have suggested that infants conceptualize the various principle-based expectations reviewed in the preceding section as *obligations* that specify how individuals should act toward others (e.g., individuals should act fairly). However, another possibility is that infants construe these expectations non-normatively, as *behavioral regularities* that capture how individuals in their social environments typically act toward others (e.g., individuals typically act fairly). How can we decide between these two possibilities? This is a challenging question to address experimentally. One approach has been to examine whether infants exhibit a negative attitude, and expect others to exhibit a negative attitude, toward individuals who do not behave as expected. In the case of fairness, for example, there is evidence that infants not only expect distributors to divide resources fairly between similar potential recipients (Buyukozer Dawkins et al., 2019; Meristo et al., 2016), but also (a) prefer fair over unfair distributors (Geraci & Surian, 2011; Lucca et al., 2018), (b) are more likely to help fair as opposed unfair distributors (Surian & Franchin, 2017), (c) detect a violation when individuals choose to reward unfair as opposed to fair distributors (Meristo & Surian, 2013), and (d) associate praise with fair distributors and admonishment with unfair distributors (DesChamps et al., 2015). Together, these results suggest that infants who observe an unfair distribution do not simply represent it as a deviation from a behavioral

regularity; rather, they view it as a norm violation, with consequences for their and others' attitude toward the wrongdoer.

To provide converging evidence that infants understand moral obligations, we have adopted a different approach focused on *virtuous* actions—namely, positive actions that are not obligatory in the situation and hence go beyond what is morally expected. In a typical experiment, two individuals, A and B, produce the same positive action toward the same protagonist, one at a time. For A, this action is obligatory (e.g., is dictated by the principle of ingroup support); for B, this action is optional. Next, infants are presented with both individuals in an implicit preferential-looking task (adapted from Hamlin et al., 2010). The rationale is that if infants look preferentially at B over A, it indicates that they understand moral obligations: (a) they recognize that A's positive action is obligatory but B's is not; (b) they infer that B is acting virtuously (e.g., out of kindness) and going beyond what is morally required in the situation; and hence (c) they evaluate B more favorably than A.

To illustrate our approach, in a series of experiments (Ting & Baillargeon, 2018b), 18-month-olds received two orientation trials, two familiarization trials, and one preferential-looking trial. In the first orientation trial, infants were introduced to two groups of puppets; there were two owls (O1 and O2) and two frogs (F1 and F2), who all spoke English and wore distinguishing accessories. In the second orientation trial, one puppet (e.g., O1) built a tower by stacking five discs of decreasing sizes, while the other puppets watched. In the two familiarization trials, O1 now needed help to complete her tower. In the *ingroup-helps* trial, O1 was initially alone building her tower and was joined midway through by O2. O1 was unable to reach the last disc, which lay across the apparatus floor from her but within O2's reach (Warneken & Tomasello, 2006). O2 helped by bringing the disc closer to O1, who then completed her tower. The *outgroup-helps* trial was identical except that F2 joined and helped O1. Finally, in the preferential-looking trial, infants saw O2 and F2 standing on either side of the apparatus; infants' looking at each puppet was coded frame-by-frame offline. Confirming prior results, infants looked equally at the ingroup- and outgroup-helps trials. During the preferential-looking trial, however, infants looked significantly longer at the outgroup helper than at the ingroup helper. Infants thus implicitly preferred the puppet who had helped out of

kindness (i.e., F2 had no obligation to help O1, so her actions provided evidence that she was kind) over the puppet who might have helped solely out of duty (i.e., O2 was obligated to help O1, so her actions provided no evidence about whether she was kind or not). This positive result was replicated in additional experiments but was eliminated when O1 did not need help and simply exchanged friendly greetings with O2 and F2; when O1 did not need help and O2 and F2 brought closer an extraneous object; and when O1 remained alone in both familiarization trials and was unable to complete her tower. Together, these results suggest two conclusions. First, infants' sociomoral expectations capture moral obligations rather than mere behavioral regularities: Here, in particular, infants viewed individuals as having an obligation to help ingroup members in need of assistance. Second, infants already appreciate virtuous actions that go beyond what is morally required.

The preceding results mirror recent findings with adults reported by McManus and Young (2019). In a series of experiments, adults were presented with vignettes about individuals who did or did not provide help to kin or strangers. Consistent with the principle of ingroup support, adults judged individuals who failed to help kin less favorably than they did individuals who failed to help strangers. Critically, and in line with our infant results above, adults also judged individuals who helped strangers more favorably than they did individuals who helped kin, presumably because they realized that whereas the latter individuals were morally obligated to help, the former individuals were not and hence went beyond what was morally expected when they showed kindness to strangers.

Moral status. How do infants identify novel individuals, endowed with the moral obligations and rights we have been discussing, such as the obligation to treat others fairly and the right to be treated fairly by others? To put it another way, on what novel entities are infants likely to confer moral status? The early sociomoral literature suggests that infants confer moral status on any novel entities they identify as *animate*. According to research on early biological reasoning (Baillargeon et al., 2016; Setoh, Wu, Baillargeon, and Gelman, 2013), a novel entity is animate if it gives evidence of having both a functioning body (e.g., it is capable of self-propulsion, which is a sign of internal energy) and a functioning mind (e.g., it interacts with its environment in an

agentive manner, which is a sign of internal control). Consistent with this analysis, studies on sensitivity to fairness in resource-distribution tasks indicate that infants expect a fair treatment for novel entities that are both self-propelled and agentive, but hold no particular expectation about the fair treatment of novel entities that are only self-propelled, only agentive, or neither self-propelled nor agentive (Buyukozer Dawkins et al., 2019; Meristo et al., 2016; Ting & Baillargeon, 2018c). For example, in recent experiments (Ting & Baillargeon, 2018c), 14-month-olds were first introduced to two colorful boxes that were devoid of any morphological similarities to humans and other animals. While an experimenter watched, the boxes gave evidence of being both self-propelled and agentive (they moved on their own and beeped contingently to each other, as though taking turns in a conversation; self-propelled/agentive condition), only self-propelled (only one box moved and beeped in each trial; self-propelled/non-agentive condition), or only agentive (the two boxes beeped contingently to each other but never moved; agentive/non-self-propelled condition). Next, the experimenter divided two identical toys between the two boxes, either fairly (*equal* event) or unfairly (*unequal* event). Infants in the self-propelled/agentive condition looked significantly longer at the unequal than at the equal event, whereas infants in the other two conditions looked equally at the events. The two boxes were thus granted moral status and, with it, the right to be treated fairly, only when they were shown to have both a functioning body and a functioning mind and were therefore perceived as animates.

The preceding results dovetail well with evidence by Weisman, Dweck, and Markman (2017) that adults' intuitive conception of mental life has three fundamental components: capacities related to having a body (e.g., initiating one's motion, experiencing hunger), capacities related to having a mind (e.g., detecting one's environment, pursuing goals), and capacities related to having a heart, in its metaphorical sense (e.g., telling right from wrong, exercising self-restraint). When infants confer moral status on a novel entity and endow it with moral obligations and rights, they presumably also endow it with the suite of mental capacities necessary to apply these obligations and rights (e.g., to determine, in a reward-dispensation situation, what is the equitable treatment for each potential recipient). If we think of granting these capacities as akin to granting a heart, it suggests that (a) infants' conception of

individuals includes a functioning heart as well as a functioning body and a functioning mind, and (b) the tripartite ontology of mental life described by Weisman and colleagues emerges early in life, is highly abstract, and supports the identification of novel, unfamiliar individuals such as faceless animate boxes.

Moral circle. We have suggested that when infants assign individuals to the same moral circle, they bring to bear a principle of harm avoidance that sets broad limits on the amount of harm the individuals can inflict on each other. But how do infants identify moral circles? At present, little is known about this issue. One possibility might be that infants initially posit a single moral circle, peopled by humans and "honorary humans" such as verbal puppets. In this view, infants would expect harm avoidance in interactions between humans, broadly defined, but would hold no particular expectation about harm avoidance in interactions (a) between humans and non-humans or (b) between non-humans. Another possibility might be that infants tend to assign different animate kinds to different moral circles. From an evolutionary perspective, such a tendency could be rooted in the never-ending battle for survival between predators and prey: Tigers and gazelles cannot belong to the same moral circle, nor can foxes and chickens, birds and worms, hawks and mice, and humans and most other animals. (The world of fiction is, of course, replete with clashes between humans and fictitious moral circles, such as the terrifying xenomorphic creatures in the "Alien" movie franchise). In this view, infants would expect harm avoidance *within* each animate kind, but would hold no expectation for harm avoidance *between* animate kinds.

Evidence for this second possibility comes from recent experiments (Jin & Baillargeon, 2019) in which 12-month-olds watched interactions between two novel animate characters with eyes and stick arms. In the same-kind condition, the characters differed only in size (e.g., a large and a small blue square); in the different-kind condition, they differed in size, color, and shape (e.g., a large green circle and a small blue square). In each condition, infants saw two events: a *harm event* in which the first character picked up a stick and hit the second character three times, and a *no-harm* event in which the first character picked up a stick and jumped next to the second character three times. Infants in the same-kind condition found the harm event unexpected, but infants in the different-kind condition did not (even though in each case

the same character was hit in the same way). These results thus provide evidence that infants expect harm avoidance between two novel animates from the same kind (e.g., two squares), but have no expectation for harm avoidance between two novel animates from different kinds (e.g., a circle and a square).

Could it be that infants always expect some degree of harm avoidance but simply have a much higher threshold for what constitutes unacceptable harm in interactions between members of different moral circles? Evidence against this suggestion comes from recent experiments on early biological reasoning (Ting, Setoh, Gelman, & Baillargeon, 2019). These experiments built on prior evidence that young infants expect animate entities to have insides (Setoh et al., 2013). To start, 8-month-olds were introduced to two novel animate entities, a closed cube and a closed cylinder, and shown that each was both self-propelled (i.e., initiated and changed its course of motion) and agentive (i.e., conversed with an experimenter using either beeps or quacks). Next, infants saw two events in which the experimenter manipulated the two entities. In the *no-harm* event, she picked up one of the entities (e.g., the cube), rotated it to show infants its closed bottom, and then tilled it from side to side. In the *harm* event, she selected the other entity (e.g., the cylinder), cut out its bottom with a knife, removed its biological-like insides in two large handfuls, rotated it toward the infants to show it was now hollow, and finally tilted it from side to side. Infants looked about equally at the two events, supporting the suggestion that even grievous harm is deemed acceptable when directed at an entity from a different moral circle. This negative result was not due to infants' inability to understand the events presented: In additional experiments, infants detected a violation when the gutted entity moved again, and this effect was eliminated if the entities were initially shown to be self-propelled but not agentive (i.e., not animate). Thus, although young infants understand that a novel animate entity's insides support its function, they detect no violation when a member of a different moral circle (in this case a human experimenter) removes these insides, causing loss of function or life. More generally, our discussions of moral status and moral circles both highlight the important role that biological animacy plays in morality, at least in the first years of life: Infants confer moral status on animate entities, and they assign different kinds of animate entities to different moral circles.

The suggestion that infants expect harm avoidance within but not across moral circles fits well with the large body of research with adults on the adverse consequences of *dehumanization*, or perceiving others as less than fully human. When individuals or groups are pushed to the edges of the human moral circle, or even excluded from it, they lose some or all of their human right to be protected from harm. This means that more harm is deemed acceptable against dehumanized individuals or groups (Bastian, Denson, & Haslam, 2013; Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014; Haslam & Loughnan, 2014; Kteily, Bruneau, Waytz, & Cotterill, 2015; Smith, 2011). To illustrate, Kteily and colleagues asked adults to locate various social groups, including their own, on a scale depicting human evolution from ape to modern human. Participants endorsed harsher treatments (e.g., less support for immigration, fewer relief donations) for groups they perceived to be less evolved or less human, and they also showed less compassionate responses to plights (e.g., acts of discrimination) suffered by these groups.

The evidence that dehumanization leads to less protection from harm has been found for a wide range of human targets, including individuals with reduced or impaired mental capacities (Capozza, di Bernardo, Falvo, Vianello, & Caio, 2016; Khamitov, Rotman, & Piazza, 2016; Martinez, Piff, Mendoza-Denton, & Hinshaw, 2011; Waytz, Gray, Epley, & Wegner, 2010). According to Waytz and colleagues (2010), for example, if someone "is seen as relatively mindless, then he or she receives diminished moral standing, and might be treated like an animal or an object . . . Those denied competence, civility, and agency come to be seen as subservient or animalistic, licensing people to contain them against their will and to rob them of human rights" (p. 386). In the early psychological-reasoning literature, there is substantial evidence that infants negatively evaluate individuals who behave irrationally: For example, infants ages 14–18 months were less likely to learn new information from, or to direct inquiries to, individuals who mislabeled or misused familiar objects, or who enthused over empty containers (Begus & Southgate, 2012; Koenig & Woodward, 2010; Poulin-Dubois, Brooker, & Polonia, 2011; Zmyj, Buttelmann, Carpenter, & Daum, 2010). Building on these results, one could ask whether infants would hold weaker expectations about irrational individuals' moral rights, such as the right to be protected from harm. Such

evidence would suggest that a disposition to dehumanize different others is already present in infancy, thus opening new avenues for intervention. (Conversely, but in the same spirit, it would also be interesting to explore what *humanization* interventions might bring infants to view harming non-human entities as unacceptable).

Moral character. The evidence reviewed in this chapter suggests that infants evaluate individuals' actions along a broad spectrum that includes (a) *forbidden* actions, which are negative actions that violate moral principles, (b) *dubious* or questionable actions, which also have a negative valence but do not unambiguously violate moral principles, (c) *neutral* actions, (d) *obligatory* actions, which are positive actions that are required by one or more moral principles, and (e) *virtuous* actions, which are positive actions that go beyond what is morally required and reveal virtues such as kindness and courage. This spectrum maps well onto the results from social-preference tasks: For example, infants have been shown to prefer individuals who produce virtuous actions over those who produce obligatory (Ting & Baillargeon, 2018b), neutral (Kanakogi et al., 2017), or dubious (Hamlin et al., 2007) actions; to prefer individuals who produce obligatory actions over those who produce forbidden actions (Buon et al., 2014; Geraci & Surian, 2011); and to prefer individuals who produce neutral actions over those who produce dubious actions (Hamlin et al., 2007).

These results raise important questions about the consequences of these evaluations, particularly with respect to issues of *moral character*. While moral principles specify how individuals *should* act, moral characters reflect how they *do* act. After all, an individual may know right from wrong but still choose to act selfishly, with little concern for others. From this perspective, an individual with a good moral character is thus one who acts morally rather than selfishly, whereas an individual with a bad moral character is one who tends to do the reverse (Tooby & Cosmides, 2010).

Do infants take individuals' actions to reveal something pervasive and stable about their moral characters, which allows predictions about how they are likely to act in new contexts? To address this question, we asked in two recent experiments whether 2-year-old toddlers who saw an individual violate a moral principle (a) would not find it unexpected if the individual then violated a different moral principle, but (b) would find it unexpected if the individual then acted generously toward others (Ting & Baillargeon,

2018a). In the first experiment, toddlers were first introduced to rabbit and dog puppets who spoke in female voices and wore distinguishing accessories (because the puppets were all capable of speech, we expected toddlers to see them as “honorary humans”). In the familiarization scenario, toddlers saw one puppet (e.g., a rabbit, R1) harm, either once or three times, a puppet from her own group (R2) or a puppet from the other group (D2). Next, in the test scenario, R1 divided two toys either fairly (*equal* event) or unfairly (*unequal* event) between two other puppets. When R1 harmed outgroup D2 only once (e.g., destroyed her drawing), infants still expected R1 to act fairly, suggesting that they did not draw broad negative inferences about her moral character (consistent with prior findings that a mild negative action toward outgroup individuals in the same moral circle is not viewed as forbidden). However, when R1 harmed outgroup D2 three times (e.g., destroyed her drawing, puzzle, and tower), or when she harmed ingroup R2 either once or three times, infants looked equally at the two events (see also Surian, Ueno, Itakura, & Meristo, 2018). Together, these results suggest that when R1 violated harm avoidance and/or ingroup support, toddlers inferred that there were broad and enduring deficiencies in her moral character, so they did not find it unexpected when she acted unfairly in a new context, demonstrating once again her lack of concern for others.

In the second experiment, toddlers first saw R1 harm R2 three times. In the test scenario, six toys were introduced for R1 and R to share. Infants detected a violation when R1 divided the toys generously (taking only one for herself and giving the other five to R3), but not when R3 did so (generous or lavish sharing is deemed acceptable among ingroup members; Jin, Bian, & Baillargeon, 2017). These and control results make clear that toddlers did not simply conclude that no predictions could be made about R1’s behavior because she did not follow moral rules. Rather, they concluded that R1 had a deficient moral character and was unlikely to act generously toward others.

Together, these results suggest four conclusions about toddlers’ evaluations of moral character. First, toddlers perceive the different moral principles as deeply inter-related and pertaining to the same domain of morality; thus, a violation of one principle affects predictions about other principles. Second, toddlers understand that the moral principles help keep self-interest in check in interactions with others; thus, an individual with a poor moral character is unlikely to act generously toward others. Third, infants

expect individuals' moral characters to be stable over time and to extend across situations. Finally, in some cases at least, a single negative action (e.g., like R1 harming her ingroup R2 once) may be sufficient to lead infants to draw negative inferences about an individual's moral character.

These results fit well with evidence that adults place a great deal of importance on individuals' moral characters and view them as stable and predictive of future actions (Bollich et al., 2016; Brambilla & Leach, 2014; Goodwin, 2015; Heiphetz, Strohminger, & Young, 2017; Strohminger & Nichols, 2014; Uhlmann, Pizarro, & Diermeier, 2015). Research with adults also shows that a single negative action may at times be sufficient to tarnish an individual's character, and that a bad impression, once formed, may require multiple positive actions to change that impression (Baumeister, Bratslavsky, Finkenauer, &Vohs, 2001; Riskey & Birnbaum, 1974; Rozin & Royzman, 2001; Skowronski & Carlston, 1992). An interesting direction for future developmental research will thus be to examine how many positive actions may be required to reverse toddlers' attribution of a deficient moral character to an individual.

4. Concluding Remarks

The evidence reviewed in this chapter suggests that at least four moral principles (fairness, harm avoidance, ingroup support, and authority) and four moral concepts (moral obligation, moral status, moral circle, and moral character) shape early sociomoral reasoning. Although a great deal of research is still needed to support these conclusions, the "first draft" of moral cognition appears to be remarkably sophisticated and to provide a rich foundation for infants' adaptation to their social worlds, both within the confines of their families and beyond.

5. References

Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How do infants reason about physical events? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 11-48). Oxford, England: Blackwell.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159–186.

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., & Bian, L. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer & P. R. Shaver

(Eds.), E. Borgida & J. A. Bargh (Assoc. Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 79–150). Washington, DC: American Psychological Association.

Bastian, B., Denson, T. F., & Haslam, N. (2013). The roles of dehumanization and moral outrage in retributive justice. *PloS ONE*, *8*(4), e61842.

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

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611-617.

Bian, L., & Baillargeon, R. (2019, March). *Toddlers and infants hold an abstract expectation of ingroup loyalty.* Paper presented at the Biennial Meeting of the Society of Research in Child Development, Baltimore, MD.

Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the National Academy of Sciences, 115*(11), 2705–2710.

Bloom, P., & Wynn, K. (2016). What develops in moral development? In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 347-364). New York, NY: Oxford University Press.

Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (2016). Eavesdropping on character: Assessing everyday moral behaviors. *Journal of Research in Personality*, *61*, 15-21.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*(4), 397-408.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429-444.

Buon, M., Jacob, P., Margules, S., Brunet, I., Dutat, M., Cabrol, D., Dupoux, E. (2014). Friend or foe? Early social evaluation of human interactions. *PloS ONE 9*(2):e88612.

Burns, M. P., & Sommerville, J. A. (2014). “I pick you”: The impact of fairness and race

on infants' selection of social partners. *Frontiers in Psychology, 5*, 93.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development, 84*, 422–428.

Buyukozer Dawkins, M., & Baillargeon, R. (2019, May). *Infants' conflicting expectations and attitudes toward the wealthy*. Paper presented at the Annual Meeting of the Human Behavior and Evolution Society, Boston, MA.

Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2017, August). *Evidence for an equity-based sense of fairness in infancy*. Paper presented at the Dartmouth Workshop on Action Understanding, Hanover, NH.

Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2019). Do infants in the first year of life expect equal resource allocations? *Frontiers in Psychology*, *10,* 116.

Buyukozer Dawkins, M., Ting, F., Stavans, M., & Baillargeon, R. (in press). Early moral cognition: A principle-based approach. To appear in D. Poeppel, G. R. Mangun, & M. S. Gazzaniga (Eds.-in-chief), *The cognitive neurosciences VI*. Cambridge, MA: MIT Press.

Capozza, D., Di Bernardo, G. A., Falvo, R., Vianello, R., & Caio, L (2016). Individuals with intellectual and developmental disabilities: Do educators assign them a fully human status? *Journal of Applied Social Psychology*, *46*(9), 497-509.

DesChamps, T. D., Eason, A. E., & Sommerville, J. A. (2016). Infants associate praise and admonishment with fair and unfair individuals. *Infancy*, *21*(4), 478-504.

Dupoux, E., & Jacob, P. (2007). Universal moral grammar: A critical appraisal. *Trends in Cognitive Sciences*, *11*(9), 373-378.

Gazes, R. P., Hampton, R. R., & Lourenco, S. F. (2017). Transitive inference of social dominance by human infants. *Developmental Science*, *20*(2), e12367.

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012-1020.

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology, 106*(4), 526-545.

Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*(1), 38-44.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55–130.

Hamlin, J. K. (2013a). Failed attempts to help and harm: Intention versus outcome n preverbal infants' social evaluations. *Cognition, 18,* 451-474.

Hamlin, J. K. (2013b). Moral judgment and action in preverbal infants and toddlers evidence for an innate moral core. *Current Directions in Psychological Science, 22,* 186–193.

Hamlin, J. K. (2014). Context-dependent social evaluation in 4.5-month-old human infants: The role of domain-general versus domain-specific processes in the development of social evaluation. *Frontiers in Psychology, 5*, 614.

Hamlin, J. K. (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. *Frontiers in Psychology, 5*, 1563.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26*, 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–559.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science, 13*, 923–929.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences, 108*, 19931–19936.

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399-423.

Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, *41*(3), 744-767.

Howard, L. H., Henderson, A. M., Carrazza, C., & Woodward, A. L. (2015). Infants' and

young children's imitation of linguistic in-group and out-group informants. *Child Development*, *86*(1), 259-275.

Jin, K., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences, 114*, 8199–8204.

Jin, K., & Baillargeon, R. (2019). Infants' expectations about harm avoidance within and between moral circles. Manuscript in preparation.

Jin, K., Bian, L., & Baillargeon, R. (2017, April). *What behaviors lead toddlers to infer an ingroup relation? The case of sharing*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Austin, TX.

Jin, K. S., Houston, J. L., Baillargeon, R., Groh, A. M., & Roisman, G. I. (2018). Young infants expect an unfamiliar adult to comfort a crying baby: Evidence from a standard violation-of-expectation task and a novel infant-triggered-video task. *Cognitive Psychology*, *102*, 1-20.

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. *Nature Human Behaviour*, *1*(2), 0037.

Khamitov, M., Rotman, J. D., & Piazza, J. (2016). Perceiving the agency of harmful agents: A test of dehumanization versus moral typecasting accounts. *Cognition*, *146*, 33-47.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*, 12577–12580.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). 'Native' objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development*, *13*(1), 67-81.

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: possible mechanisms. *Developmental Psychology*, *46*(4), 815.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), 901-931.&

Liberman, Z., Woodward, A. L., Sullivan, K. R., Kinzler, K. D. (2016). Early emerging

system for reasoning about the social nature of food. *Proceedings of the National Academy of Sciences*,*113*, 9480–9485.

Lucca, K., Pospisil, J., & Sommerville, J. A. (2018). Fairness informs social decision making in infancy. *PloS ONE*, *13*(2), e0192848.

Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, *115*(38), E8835-E8843.

Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology, 54*(8), 1445-1455.

Martinez, A. G., Piff, P. K., Mendoza-Denton, R., & Hinshaw, S. P. (2011). The power of a label: Mental illness diagnoses, ascribed humanity, and social rejection. *Journal of Social and Clinical Psychology*, *30*(1), 1-23.

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, *109*(18), 6862-6867.

McManus, R., & Young, L. (2019, May). *What we owe to family: The impact of social obligations on moral judgment*. Paper presented at the Annual Meeting of the Human Behavior and Evolution Society, Boston, MA.

Meristo, M., Strid, K., & Surian, L. (2016). Preverbal infants' ability to encode the outcome of distributive actions. *Infancy*, *21*(3), 353-372.

Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition*, *129*(1), 102-113.

Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, *34*(2), 303-309.

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences, 110*, 3965–3972.

Premack, D., & Premack, A. J. (1997). Infants attribute value± to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, *9*(6), 848-856.

Pun, A., Birch, S. A., & Baron, A. S. (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences*, *113*(9), 2376-2381.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*, 57–75.

Rhodes, M. (2012). Naïve theories of social groups. *Child Development*, *83*(6), 1900-1916.

Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. M. (2015). Infants' use of social partnerships to predict behavior. *Developmental Science*, *18*(6), 909-916.

Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology, 103*, 171-173.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296-320.

Schmidt, M. F., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PloS ONE*, *6*(10), e23223.

Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences*, *110*(40), 15937-15942.

Shutts, K., Kinzler, K. D., McKee, C. B., & Spelke, E. S. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development, 10*, 1–17.

Shweder, R. A., Much, N. C., Mahapatra, M. & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York, NY: Routledge.

Skowronski, J. J., & Carlston, D. E. (1992). Caught in the act: When impressions based on highly diagnostic behaviours are resistant to contradiction. *European Journal of Social Psychology*, *22*(5), 435-452.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23*, 196–204.

Smith, D.L. (2011). *Less than human: Why we demean, enslave, and exterminate others*. New York, NY: St. Martin's Press.

Spokes, A. C., & Spelke, E. S. (2017). The cradle of social knowledge: Infants' reasoning about caregiving and affiliation. *Cognition, 159*, 102–116.

Stavans, M., & Baillargeon, R. (in press). Infants expect leaders to right wrongs. *Proceedings of the National Academy of Sciences*.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*, 159-171.

Surian, L., & Franchin, L. (2017). Toddlers selectively help fair agents. *Frontiers in Psychology*, *8*, 944.

Surian, L., Ueno, M., Itakura, S., & Meristo, M. (2018). Do infants attribute moral traits? Fourteen-month-olds' expectations of fairness are affected by agents' antisocial actions. *Frontiers in Psychology*, *9,* 1649.

Thomsen, L., Frankenhuis, W., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science, 331*, 477–480.

Ting, F., & Baillargeon, R. (2018a, July). *2-year-old toddlers make broad inferences about moral characters*. Paper presented at the Biennial International Congress on Infant Studies, Philadelphia, PA.

Ting, F., & Baillargeon, R. (2018b, July). *Do infants understand moral obligations?* Paper presented at the Biennial International Congress on Infant Studies, Philadelphia, PA.

Ting, F., & Baillargeon, R. (2018c, July). *14-month-olds use animacy cues to determine who has moral rights*. Paper presented at the Annual Meeting of the Society for Philosophy and Psychology, Ann Arbor, MI.

Ting, F., He, Z., & Baillargeon, R. (2019a). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences*, *116*(13), 6025-6034.

Ting, F., He, Z., & Baillargeon, R. (2019b, March). *Group membership modulates early expectations about retaliatory harm*. Paper presented at the Biennial Meeting of the Society of Research in Child Development, Baltimore, MD.

Ting, F., Setoh, P., Gelman, R., & Baillargeon, R. (2019). Infants understand that animates cannot function without their insides. Manuscript in preparation.

Tooby, J., and Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary & comparative perspectives* (pp. 191-234). New York, NY: Palgrave MacMillan.

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

Wang, Y., & Henderson, A. M. (2018). Just rewards: 17-month-old infants expect agents to take resources according to the principles of distributive justice. *Journal of Experimental Child Psychology, 172*, 25-40.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301-1303.

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

Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, *114*, 11374-11379.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208-220.