Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core

J. Kiley Hamlin[1]

2472 Words

In Press, *Current Directions in Psychological Science*

Please do not cite or quote without permission

Abstract

Although developmental psychologists traditionally explore morality from a learning and development perspective, some aspects of the human moral sense may be built-in, having evolved to sustain collective action and cooperation as required for successful group living. This paper reviews a recent body of research with infants and toddlers, demonstrating surprisingly sophisticated and flexible moral behavior and evaluation in a preverbal population whose opportunity for moral learning is limited at best. Although itself in its infancy, this work supports theoretical claims that human morality is a core aspect of human nature.

Ask a group of 5-year-olds to tell you a story, and they will likely regale you with tales of virtuous princesses and evil stepmothers, of superheroes and supervillians, or perhaps of the time someone misbehaved and was sent to bed without dessert. This tendency to see certain actions and individuals as right, good, and deserving of reward, and others as wrong, bad, and deserving of punishment – generally referred to as a "moral sense" – is often considered a defining feature of what it means to be human. Indeed, although there are considerable cross-cultural differences regarding which actions are considered obligatory, permissible, or forbidden, and a wide variety of prescribed responses to violations across societies, all normally developing adults in all cultures have the basic notion that some things are right and others are wrong (Brown, 1991). But where does this moral sense come from?

Traditional accounts: Morality is learned

Traditionally, psychologists have probed moral origins from a learning and development perspective (e.g., Kohlberg, 1969; Piaget, 1932; see Killen & Smetana, 2006, for reviews). These accounts posit that infants start out as fundamentally different moral creatures from adults – either *a*moral (possessing no moral sense) or *im*moral (possessing a moral sense that opposes adults', perhaps due to selfishness or cognitive limitations) – and acquire a mature moral sense over time through various developmental processes. Indeed, extensive empirical work demonstrates that as children become increasingly other-focused, experienced, socialized, and cognitively skilled, they show corresponding improvements in morally-relevant cognitions and evaluations (see Killen & Smetana, 2006);

furthermore, cross-cultural variation in the moral sense in adulthood can only have resulted from varied input during development.

However, development and cultural variation do not themselves preclude the existence of innate capacities. As has been argued for humans' "core" understanding of certain conceptual domains including objects, number, geometry, agents, and groups (see Spelke, 1994; Spelke & Kinzler, 2007), there may be aspects of morality that emerge in the absence of specific experiences, such as (amongst others) being helped or harmed in particular ways/situations, observing others be similarly helped or harmed, or being explicitly taught which acts are right and which are wrong. This "moral core" would remain essentially intact throughout the lifespan, constraining how subsequent experiences and advances in other domains influence moral development. Indeed, scholarly interest in such an innate moral core has blossomed in recent years (e.g., Haidt, & Joseph, 2007; Hamlin, 2010; Joyce, 2006; Katz, 2000; Mikhail, 2011; Premack & Premack, 1994).

Morality for cooperation

Evolutionary biologists, anthropologists, and primatologists are increasingly exploring the integral nature of morality to successful group living (see Alexander, 1987; Cosmides & Tooby, 1992; de Waal, 2006; Henrich & Henrich, 2007; Joyce, 2006; Katz, 2000; Price, Cosmides, & Tooby, 2002). Though the details vary, these functional accounts generally posit that our moral sense evolved in order to sustain collective action and cooperation – which lead to great mutual gain but sometimes require personal sacrifice – within groups of unrelated individuals. Cooperative systems require (at least) three abilities/propensities, all relevant to the moral sense: (1) moral *goodness*: the tendency to feel concern for and help others despite personal costs, presumably rooted in empathetic processes; (2) moral *evaluation*: the ability to identify and dislike those who are uncooperative/unempathetic/unhelpful, or who are likely to be in the future, requiring an ability to analyze others' social behaviors; and (3) moral *retribution*:

the tendency to carry out, or to support, punishment of those who misbehave, which could involve emotional processes, behavioral analyses, or both. Because humans are arguably the most cooperative, empathetic, and altruistic species on earth (e.g., Tomasello, 2009), we should also be most evaluative and retributive.

Requirement (1): Moral goodness

A growing body of developmental data suggests Requirement (1), moral goodness, is present in infancy and rooted in empathy. From birth, newborns show rudimentary emotional reactions to others' suffering (Martin & Clark 1982; Sagi & Hoffman, 1976). As soon as they are physically able, infants begin to supplement these emotional responses with a variety of active prosocial behaviors, including comforting those in distress, helping others achieve goals, informing others of things they should know, and sharing their own resources (see Dunfield, Kuhlmeier, O'Connell, & Kelley, 2010; Eisenberg, Fabes, & Spinrad, 2006; Warneken & Tomasello, 2009, for reviews). Although active prosocial behaviors emerge after birth, they are unlikely wholly the result of brute socialization: they occur spontaneously, are present in our primate relatives, and are intrinsically motivated (see also Aknin, Hamlin, & Dunn, 2012; Hepach, Vaish, & Tomasello, 2012).

Requirement (2): Moral understanding and evaluation

If arguments regarding the co-evolution of cooperation and morality are correct, early prosociality should be accompanied by capacities for moral understanding and evaluation. Specifically, these capacities should relate to behaviors that are relevant to cooperation, such as helping/assisting, hindering/preventing, giving/receiving/taking, and fairness/inequity. Consistent with this requirement, a growing body of work suggests that preverbal infants understand several types of cooperative behaviors that occur between third parties. By the end of their first year, infants grasp that agents work together toward shared goals (Henderson & Woodward, 2011), categorize goal-helping as positive and goal-

hindering as negative (Premack & Premack, 1997), and understand how being helped and hindered influences one's social preferences (e.g., Kuhlmeier, Wynn, & Bloom, 2003). By the middle of the second year, they expect individuals to treat others fairly (equally dividing resources; e.g., Geraci & Surian, 2011), as long as everyone has contributed equally to the cooperative endeavor and/or is in the same group (Sloane, Baillargeon, & Premack, 2012).

In addition to understanding cooperatively-relevant behavior, infants positively evaluate those who cooperate, and negatively evaluate those who do not. The first studies of this nature were reported by Hamlin and colleagues (e.g., Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; 2010). Infants viewed live "Morality Plays," in which a "Protagonist" puppet (P) tried, but failed, to achieve a goal. Several goal-failure scenarios were tested, including a wooden, googly-eyed puppet trying to reach the top of a hill (adapted form Kuhlmeier et al., 2003), an animal puppet trying to open a box containing a toy, and an animal puppet requesting the return of a ball it accidentally dropped (stimuli depicted in Figure 1, videos available at http://www.cic.psych.ubc.ca/Current_Directions_Media.html). On alternating events, a "Helper" facilitated P's goal (bumped P up the hill/opened the box/returned P's ball), and a "Hinderer" prevented P's goal (bumped P down the hill/slammed the box closed/stole P's ball). Infants' preference for the Helper versus the Hinderer was determined by reach behavior (Figure 2). Remarkably, from as early as they can physically reach for objects (by about 4.5 months), between 75% and 100% of infants in *every single study* preferred the Helper to the Hinderer. Follow-up experiments measured the preferential attention of 3-month-olds who cannot yet reach, and discovered that infants' evaluations are initially driven by an aversion to Hinderers (HW, 2011; HWB, 2010). By 6 months, infants' choice behavior indicates that they both positively evaluate Helpers and negatively evaluate Hinderers (HWB, 2007).

Figure 1: Morality Plays shown to infants. Purple boxes represent the Hill scenario used in HWB 2007; 2010. Red boxes represent the Box scenario used in HW, 2011

(also Hamlin, in prep; under review; HWBM 2011). Blue boxes represent the Ball scenario used in HW, 2011 (also Hamlin, in prep; HWBM 2011; HMLW, in press).

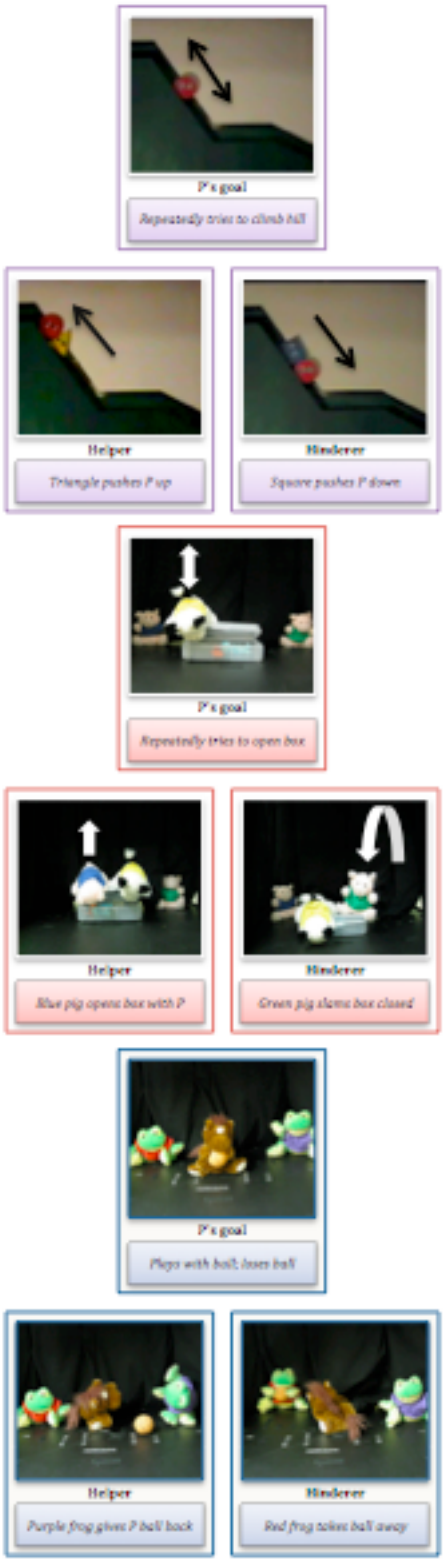
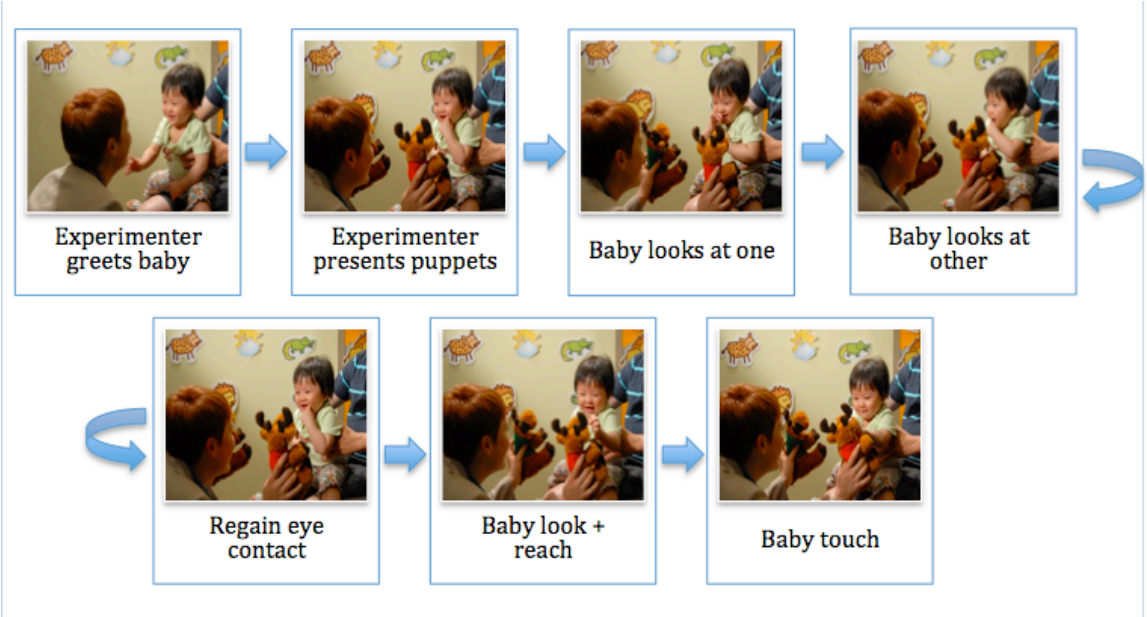


Figure 2: Choice procedure utilized with infants 4.5 months of age and older.



Evaluating low-level cues?

Infants' preferences are consistent across several puppet types, goal scenarios and preference methodologies; thus, it seems unlikely that they stem from particular perceptual cues in the stimuli (although this must be ruled out for each case; see Scarf, Imuta, Colombo, & Hayne, 2012, and response by Hamlin, Wynn, & Bloom, 2012). To ensure that perceptual cues could not underlie infants' preferences for characters in the box and ball scenarios, Hamlin & Wynn (2011) showed age-matched infants "control" plays, in which an inanimate claw, rather than an animate puppet, tugged on a box lid or bounced a ball. Identical puppets then directed the exact same actions toward the claw as they had directed toward the Protagonist during "test" plays (opened/slammed box or gave/stole ball); yet, infants did not prefer box-opening or ball-giving puppets. This result is consistent with empirical work suggesting that infants do not attribute unfulfilled goals to non-

agents such as claws (e.g., Hamlin, Newman, & Wynn, 2009; Meltzoff, 1995): because only agents have goals, puppets' actions toward non-agents are neither prosocial nor antisocial. Like adults do when making moral judgments, then, infants evaluate Helpers and Hinderers for their *goal-relevant, social* behaviors, rather than for performing some specific physical act or causing some outcome.

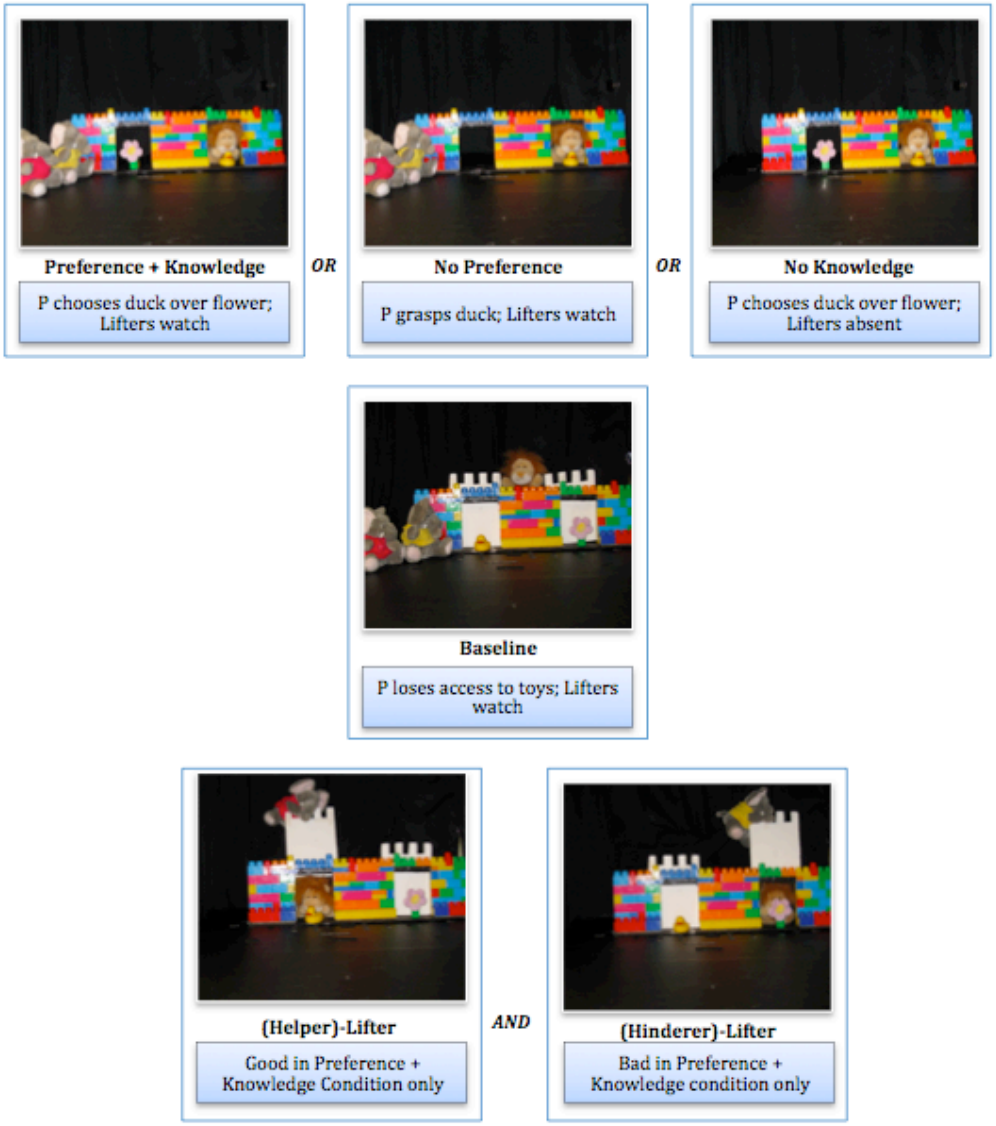
Evaluating mental states?

When determining moral responsibility, adults consider not only whether an individual facilitated or blocked a goal, but the *mental states* that drove the action. Indeed, because we sometimes act in error (e.g., unintentionally, naïvely), and because sometimes prosocial and antisocial behaviors are physically identical, mental states are critical for accurately understanding what individuals are really like, and how they will likely behave in the future – in this case, cooperatively or not. For example, consider the act of giving chocolate. If a chocolate-giver is unaware that a potential recipient hates chocolate, s/he can hardly be blamed for giving it: the giver might have been trying to be prosocial, and one should expect him or her to cooperate during future interactions. If, on the other hand, the giver *knows* the recipient hates chocolate but gives it anyway, this same chocolate-giving act is antisocial, and one might wish to avoid this individual in the future. To accurately evaluate a giving behavior, then, one must read the mental states of all parties: the giver's intention (mental state 1) toward the recipient, based on the giver's knowledge (mental state 2) of the recipient's preference (mental state 3).

Do infants consider mental states when evaluating others for their helpful and unhelpful acts? Although failure to take mental states into account has long been considered fundamental to young children's moral immaturity (Piaget, 1932), several recent studies demonstrate that infants successfully analyze actions that involve higher-order mental states such as ignorance and false belief, if tested non-verbally (e.g., Kovacs, Teglas, & Endress, 2010; Luo, 2011; Onishi & Baillargeon, 2005; see also Perner & Ruffman, 2005). Hamlin, Ullman, Tenenbaum, Goodman, &

Baker (in press) explored whether 10-month-olds incorporate such mental-state reading abilities into their social evaluations. In a Preference + Knowledge condition (see Figure 3), P exhibited a preference for one of two toys by repeatedly grasping one toy and not the other. Two additional "Lifter" puppets sat onstage facing P, implying they "knew" P's preference. P then lost access to both toys, and one Lifter acted (lifted a door) to allow P to reach the toy he had chosen (Helper), and the other Lifter acted identically (lifted a door) to allow P to reach the toy he had avoided (Hinderer). Despite their physically-identical actions, infants strongly preferred the Helper-Lifter to the Hinderer-Lifter. Two additional conditions systematically removed either P's, or the Lifters', mental states: In the 'No Preference' condition, only one toy was present during P's grasps and no preference was attributable to him; in the 'Ignorance' condition, the Lifters were off-stage during P's initial toy-grasps and were therefore ignorant of P's preference. Infants chose Lifters equally in each of these conditions, indicating that in the Preference + Knowledge condition they did not simply prefer the puppet who happened to facilitate P's goal, or who allowed P to grasp the toy he'd grasped before: infants evaluated Lifters based on their prosocial and antisocial mental states.

Figure 3: Morality Play shown to infants in HUTGB, in press.



Critically, these results demonstrate that infants' evaluations are not limited to simple heuristics like "causing-someone-to-do-X=good", "former-action-facilitating=good", or even "intention-facilitating=good"; these heuristics applied to conditions in which infants failed to distinguish characters. Instead, like adults, infants evaluate others as good and bad mentalistically: good guys are those who knowingly and intentionally facilitate a third party's goal; bad guys are those who knowingly and intentionally block it. These results add to a growing body of evidence that infants can appreciate the critical role that mental states play in others' actions, if they are tested using developmentally-appropriate methodologies

(e.g. Kovacs, Teglas, & Endress, 2010; Luo, 2011; Onishi & Baillargeon, 2005). Indeed, as prosocial and antisocial mental states should predict others' likelihood to cooperate in the future, such a mentalistic evaluative capacity is crucial for supporting cooperation on a group level, allowing individuals to selectively cooperate with prosocial and avoid antisocial others.

Requirement (3): Retribution

The third requirement for the evolution of cooperation is moral retribution: there must be a way to punish those who cheat the system and to deter potential free-riders, lest the cooperative system collapse. In other words, moral evaluations need teeth. On the surface, retaliation appears to conflict with the tendency to negatively evaluate those who intentionally block others' goals (Requirement (2)): punishment is inherently goal-blocking. Yet, adults easily solve this conflict by interpreting actions in terms of their global, rather than their local, value (see Heider, 1958): in some cases intentionally and knowingly blocking others' goals is considered the "right" thing to do, whether it be to punish bad behavior, keep vulnerable individuals safe, be loyal to one's allies, etc. Thus, a core system for moral goodness should allow for goal-preventing behaviors, and a core moral evaluation tendency should excuse, or even celebrate, those who perform justified goal-blocking acts intentionally.

Before they are 2, toddlers direct their own antisocial behaviors appropriately, selectively taking resources from someone who previously hindered a third party (Hamlin, Wynn, Bloom, & Mahajan, 2011). In addition, rather than always liking those who intentionally facilitate goals, preverbal infants prefer appropriately-antisocial characters (who harm Hinderers) over inappropriately-prosocial ones (who help Hinderers; HWBM, 2011). This effect has been demonstrated in infants as young as 5 months (Hamlin, in prep), many of whom have presumably *never seen* an act of punishment (the majority of subjects are first borns), much less been urged to positively evaluate one.

Two possible mechanisms could underlie infants' tendency to prefer those who harm Hinderers: they might believe Hinderers "deserve punishment," and thus like those who perform it; alternatively, infants might positively evaluate those who hinder antisocial others as "enemies of their enemy." Either mechanism is flexible enough to support cooperation on a grand scale; however, an "enemy of my enemy" mechanism is presumably more beneficial: it can be applied to anyone who is likely to cause harm, including Hinderers, but also out-group members, friends of enemies, etc. Indeed, infants also prefer those who hinder (versus help) puppets who are *different from them* (Hamlin, Mahajan, Liberman, & Wynn, in press), suggesting the "enemy" mechanism may underlie their nuanced evaluations.

Conclusion

In sum, recent developmental research supports the claim that at least some aspects of human morality are innate. From extremely early in life, human infants show morally-relevant motivations and evaluations, ones that are mentalistic, nuanced, and that do not appear to stem from socialization or morally-specific experience. Indeed, these early tendencies are far from shallow, mechanical predispositions to behave well, or knee-jerk reactions to particular states of the world: infants' moral inclinations are sophisticated, flexible, and surprisingly consistent with adults', incorporating aspects of moral goodness, evaluation, and retaliation. This research supports theorizing on the co-evolution of cooperation and morality, and suggests that morality is a core aspect of human nature. Future research should focus on how this early-emerging moral core combines with experience and other developmental mechanisms to create a culturally-specific, adult moral sense.

Author Notes

[1] Address correspondence to: J. Kiley Hamlin, Department of Psychology, University of British Columbia, kiley.hamlin@psych.ubc.ca

The author declares no conflicts of interest.

I would like to thank my advisors and collaborators Karen Wynn and Paul Bloom, as well as Elizabeth Dunn, Janine Gellerman, David Tannenbaum & Jessica Tracy for their helpful comments.

References

Aknin, L.B., Hamlin, J.K., Dunn, E.W. (2012). Giving leads to happiness in young children. *PLoS ONE 7(6):* e39211. doi:10.1371/journal.pone.0039211.

Alexander, R.D. (1987). *The Biology of Moral Systems.* New Jersey: Transaction Publishers.

Brown, D.E. (1991). *Human Universals*. New York: McGraw Hill.

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, J., Cosmides, L. & Tooby, J., (Eds). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

deWaal, F. (2006). *Primates and Philosophers: How morality evolved.* Princeton: Princeton University Press.

Dunfield, K.A., Kuhlmeier, V.A., O'Connell, L., & Kelley, E. (2011). Examining the diversity of prosocial behavior: helping, sharing, and comforting in infancy. *Infancy*, *16,* 227-247.

Eisenberg, N., Fabes, R. A., & Spinrad, T. (2006). Prosocial Development. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 646–718). Hoboken, NJ: John Wiley & Sons.

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.*

Haidt, J. & Joseph, C. (2007). The moral mind: How five sets of moral intuitions guide the development of many culturally-specific virtues, and perhaps even modules. Pp 367-392. In P. Carutthers, S. Laurence, & S. Stich (Eds.) *The Innate Mind, Volume 3.* Oxford: Oxford University Press.

Hamlin, J.K. (2010). Social evaluation by preverbal infants. Unpublished dissertation, Yale University.

Hamlin, J.K. (in prep). When antisocial others are good: Data from 5- and 23-month-olds.

Hamlin, J.K., Mahajan, N., Liberman, Z., & Wynn, K. (in press). Not like me = bad:] Infants prefer those who harm dissimilar others. *Psychological Science.*

Hamlin, J.K., Newman, G., & Wynn, K. (2009). 8-month-olds infer unfulfilled goals, despite contrary physical evidence. *Infancy, 14(5),* 579-590.

Hamlin, J.K., Ullman, T., Tenenbaum, J.B., Goodman, N., & Baker, C. (in press). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science.*

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929.

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

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Henderson, A.M.E., & Woodward, A. L. (2011). Let's work together: What do infants understand about collaborative goals? *Cognition, 121,* 12-21.

Henrich, N., & Henrich, J. (2007). *Why humans cooperate: A cultural and evolutionary explanation.* Oxford: Oxford University Press.

Hepach, R., Vaish, A., & Tomasello, M. (2012). A new look at children's prosocial motivation. *Infancy,* DOI: 10.1111/j.1532-7078.2012.00130.x.

Joyce, R. (2006). *The Evolution of Morality.* Cambridge: MIT Press.

Katz, L.D. (2000). *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* Thorverton, UK: Imprint Academic.

Killen, M. & Smetana, J. (2006). *Handbook of Moral Development*. New Jersey: Lawrence Erlbaum Associates, Inc.

Kohlberg, L. (1969) Stages in the Development of Moral Thought and Action, New York: Holt, Rinehart & Winston.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402.

Luo, Y. (2011). Do 10-month-olds understand others' false beliefs? *Cognition, 121(3),*

289-298.

Martin, G.B., & Clark, R.D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology, 18*, 3-9.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838–850.

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

Onishi, K. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258.

Piaget, J. (1932). *The Moral Judgment of the Child*. London: Kegan Paul, Trench, Trubner and Co.

Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308(5719)*, 214-216.

Premack, D. & Premack, A.J. (1994). Moral belief: Form versus content. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press.

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

Price, M.E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior*, 23, *203-231.*

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology*, *12*(2), 175–176.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23(2),* 196-204.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition, 50*, 431-445.

Spelke, E. & Kinzler, K. (2007). Core Knowledge. *Developmental Science, 10(1)*, 89-96.

Tomasello, M. (2009). *Why We Cooperate.* Cambridge: MIT Press.

Warneken, F. & Tomasello, M. (2009). Varieties of altruism in children and chimpanzees. *Trends in Cognitive Science, 13,* 397-402.

Recommended Readings

Alexander, R.D. (1987). (See References). One of the original works on the evolution of morality.

Killen, M. & Smetana, J. (2006). (See References). A comprehensive account of classic moral development research.

Hamlin, J.K., Wynn, K., & Bloom, P. (2007). (See References). Presents the first studies to investigate preverbal infants' moral evaluations using the "Morality Play" methodology.

Krebs, D. (2008). Morality: An evolutionary perspective. *Perspectives on Psychological Science, 3(3)*, 149-172.

A recent review of evolutionary theories of morality.

Warneken, F., & Tomasello, M. (2009). (See References). A review of recent work on early prosocial behaviors; includes primate references.