

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

The Study of Early Social Evaluation: Contextualizing Failures to Replicate and Looking Forward

Permalink

<https://escholarship.org/uc/item/9wq029xp>

Journal

Open Mind, 9

ISSN

2470-2986

Authors

Woo, Brandon
Laha, Anushka
Chen, Amelia
[et al.](#)

Publication Date

2025-08-22

DOI

10.1162/opmi.a.29

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

This paper is currently undergoing peer review.

The Study of Early Social Evaluation:

Contextualizing Failures to Replicate and Looking Forward

Brandon Woo, Anushka Laha, Amelia Chen, and Carly Wolf

Department of Psychological and Brain Sciences, University of California, Santa Barbara

Correspondence should be addressed to Brandon Woo (bmwoo@ucsb.edu).

ORCID ID

Brandon M. Woo: <https://orcid.org/0000-0002-8639-2919>

Acknowledgements

We thank Kiley Hamlin and Francis Yuen for helpful discussion.

Abstract

In classic research, Hamlin et al. (2007) found that infants preferentially reach to agents who help others over agents who hinder others. These early findings provided evidence that infants may engage in social evaluation. There have now been multiple failures to replicate the original findings, and these failed replication attempts have called into question whether infants engage in social evaluation. In the present paper, we contextualize these failed replication attempts in the broader literature. We review evidence that infants and toddlers evaluate a range of social behaviors, including acts of physical harm, resource distribution, comforting, imitation, and dominance. We also review evidence that infants and toddlers have demonstrated other behaviors, beyond preferential reaching, that suggest that they engage in social evaluation. The larger body of evidence therefore supports the possibility that early social evaluation is a real phenomenon. We discuss how cognitive scientists can better study early social evaluation moving forward.

Keywords: cognitive development, infancy, social evaluation, methods

The Study of Early Social Evaluation:

Contextualizing Failures to Replicate and Looking Forward

As adults, we can look at other people's interactions, and we can make multiple inferences: what people want, how much they like each other, whether they are prosocial or antisocial, and more. How do we come to this knowledge? Scholars have long studied and debated the development of social reasoning. While some philosophers have argued that human knowledge is based on our experiences (e.g., Hobbes, 1651/1696; Hume, 1739/2003; Locke, 1689/1847), others have instead argued that infants may have early-emerging cognitive capacities (Chan, 2002), and others still have argued that these possibilities are not mutually exclusive and that human knowledge may be based on some combination of early-emerging capacities and experience (Dewey, 1922/2002). Developmental psychologists have contributed evidence to push these debates forward. The present paper examines the evidence for the possibility that human infants engage in social evaluation.

For a long time, studies of social cognitive development were largely restricted to studies of children who could respond verbally to questions (Kohlberg & Kramer, 1969; Piaget, 1932/1965; Turiel, 1983). Then, in the late 20th century and in the 2000s, everything changed: Researchers began examining the nonverbal behaviors of infants to try to determine what they knew of the social world.

In a high-profile study, published in *Nature*, Hamlin et al. (2007) presented 6- and 10-month-old infants with a protagonist agent (a circle with googly eyes) who repeatedly tried to climb a hill. In alternating trials, one agent (e.g., a triangle) pushed the protagonist up the hill (i.e., helping), whereas another agent (e.g., a square) instead pushed the protagonist down the hill (i.e., hindering) (Fig. 1A). After infants had habituated to these events, the researchers presented

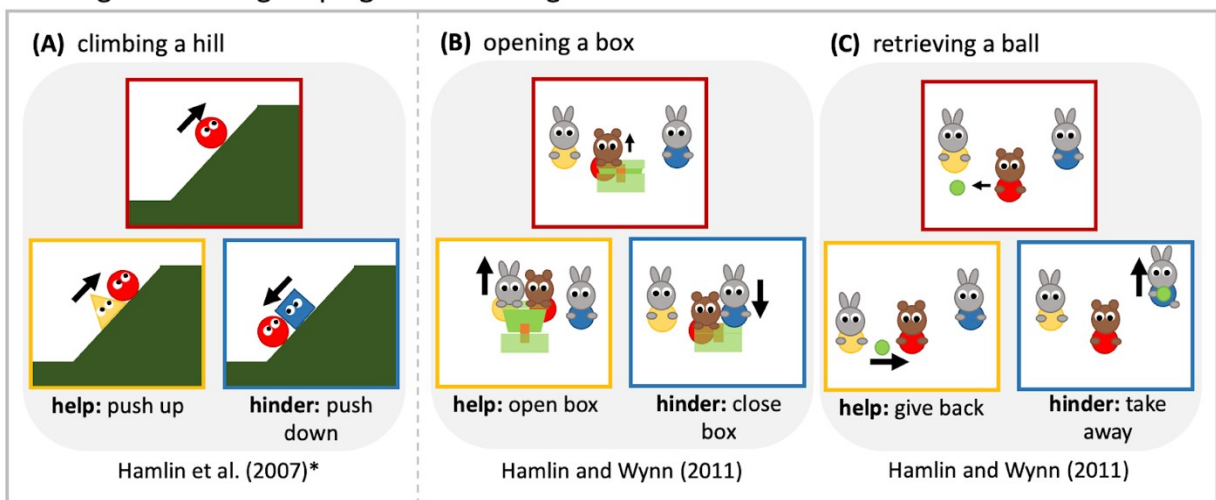
infants with the helper and the hinderer characters, and the infants preferentially reached for the helper over the hinderer. Additionally, in a further study, infants preferentially reached for the helper over a neutral character who had not interacted with the protagonist, and for a neutral character over a hinderer. Together, these findings were taken as evidence of social evaluation in infancy: that infants evaluated helpers and hinderers. The paper has been highly impactful, having been cited over 2000 times. The sample sizes of this paper were relatively small, with some of the studies having samples of 8 or 16 infants. Since 2007, a number of studies have shed light on the breadth and depth of early social evaluation. Together, this body of research suggested that infants are not blank slates, but that they may instead have a moral core: early-emerging abilities to make sense of others' prosocial and antisocial behavior (Hamlin, 2013; Rottman & Young, 2015; Ting et al., 2019; Woo et al., 2022; see also, Sommerville, 2024).

However, there have since been multiple failures to replicate the original findings and related findings (e.g., Cowell & Decety, 2015; Schlingloff et al., 2020; see also, Woo et al., 2024). In a recent paper, the ManyBabies Consortium, an international group of researchers (including the original first author of the 2007 paper) tested 302 infants, from 5 to 10 months of age (Lucca, Yuen, et al., 2025). They used stimuli based on those of Hamlin et al. (2007), in which a protagonist attempted to climb a hill. Would these 302 infants reach for the helper or the hinderer? The ManyBabies Consortium failed to replicate the finding that infants prefer helpers over hinderers; infants instead appeared to reach at chance levels, with only 49% of infants preferring the helper over the hinderer. The ManyBabies Consortium's failed replication and other failed replication attempts have called into question not only the original finding, but the study of early social evaluation more broadly (Melis & Raihani, 2023; Schmidt & Rakoczy,

2023). These failures to replicate may reflect that early social evaluation is fragile, that there could be greater individual differences than originally thought, or both.

In the present paper, we attempt to make sense of these failures to replicate, contextualizing them in the broader literature and discussing ways forward for the study of early social evaluation. First, we will review evidence that infants and toddlers evaluate a wide range of social behaviors, beyond helping and hindering (see schematics of stimuli in Fig. 2 and sample sizes and effect sizes in Fig. 3). Second, we will review evidence that infants and toddlers have demonstrated other behaviors, beyond preferential reaching, that suggest that they engage in social evaluation.

Paradigms involving helping and hindering



*the focus of most efforts to replicate evidence of early social evaluation

Figure 1. Paradigms involving helping and hindering. Across panels, the protagonist has a goal, and is either helped or hindered. This figure was adapted with permission from Woo et al. (2022).

Infants and Toddlers Evaluate a Wide Range of Social Behavior

Although the ManyBabies Consortium failed to replicate the finding that infants prefer helpers over hinderers, as the consortium themselves noted, they only studied one paradigm involving helping and hindering: the paradigm involving a protagonist trying to climb a hill. Researchers have now uncovered evidence that infants and toddlers appear to evaluate other kinds of helping and hindering. Additionally, there is evidence that infants and toddlers evaluate prosocial and antisocial behavior more broadly (e.g., acts of physical harm, resource distribution, and comforting), as well as other social behaviors, like imitation and acts of dominance. The kinds of agents have varied across these studies, and they have included shapes (e.g., Kanakogi et al., 2013; Thomas et al., 2018), puppets (e.g., Hamlin & Wynn, 2011), cartoon characters (e.g., Geraci & Surian, 2011), and human agents (e.g., Buon et al., 2014). The broader literature therefore provides evidence that infants and toddlers are capable of social evaluation.

Additional Forms of Helping and Hindering. Following Hamlin et al.'s (2007) original work, Hamlin and colleagues examined whether infants evaluate other forms of helping. Hamlin and Wynn (2011) devised two new paradigms. In the first, a protagonist attempted to open a box, a helper opened the box, and a hinderer slammed the box lid shut (Fig. 1B). In the second, a protagonist lost a ball, a helper returned the ball, and a hinderer took the ball away (Fig. 1C). In both of the original experiments, infants preferred the helper over the hinderer (see Fig. 3). Specifically, 5- and 9-month-old infants preferentially reached for the helper over the hinderer in the box paradigm, 5-month-old infants preferentially reached for the helper over the hinderer in the ball paradigm, and 3-month-old infants preferentially looked to the helper over the hinderer in the ball paradigm (for evidence in other paradigms that infants and toddlers evaluate agents

who do or do not help others obtain resources, see Hamlin et al., 2013; Van de Vondervoort et al., 2017; Woo et al., 2017). We note that these early studies that developed additional paradigms to probe infants' evaluations of helping and hindering came exclusively from Hamlin and colleagues.

There have since been attempts to replicate and build on both the box and ball paradigms. There has been relatively more mixed evidence concerning replication attempts of the box paradigm, with some success (e.g., Steckler et al., 2017, from Hamlin's lab; see also, Woo & Spelke, 2023b, 2023a) and some failures to replicate it. In two experiments, Salvadori et al. (2015), an independent group of researchers, attempted to replicate the box paradigm findings and found that 9-month-old infants did not prefer the helper over the hinderer above chance (see Fig. 3).

In contrast, there has not been the same level of mixed evidence for the finding that infants prefer agents who return a lost ball over agents who do not. For example, in one condition of an experiment, Scola et al. (2015), another independent group of researchers, attempted to replicate the ball paradigm findings and found that infants and toddlers preferred the helper over the hinderer (see Fig. 3). In a meta-analysis of all the preferential reaching data to date at the time, Margoni and Surian (2018) found that across labs, preferences for helpers were higher in the ball paradigm than in the box paradigm and the hill paradigm. Margoni and Surian also noted, however, that at the time, most of the evidence for evaluations of helping and hindering had come from Hamlin and colleagues.

Additional Prosocial and Antisocial Behaviors. Building on the research of Hamlin et al. (2007), other researchers have since explored other prosocial and antisocial behaviors that

infants and toddlers may evaluate (Woo et al., 2022). There is evidence that infants and toddlers evaluate acts of harm, acts of resource distribution, and comforting behavior.

There is evidence that infants and toddlers evaluate agents who engage in:

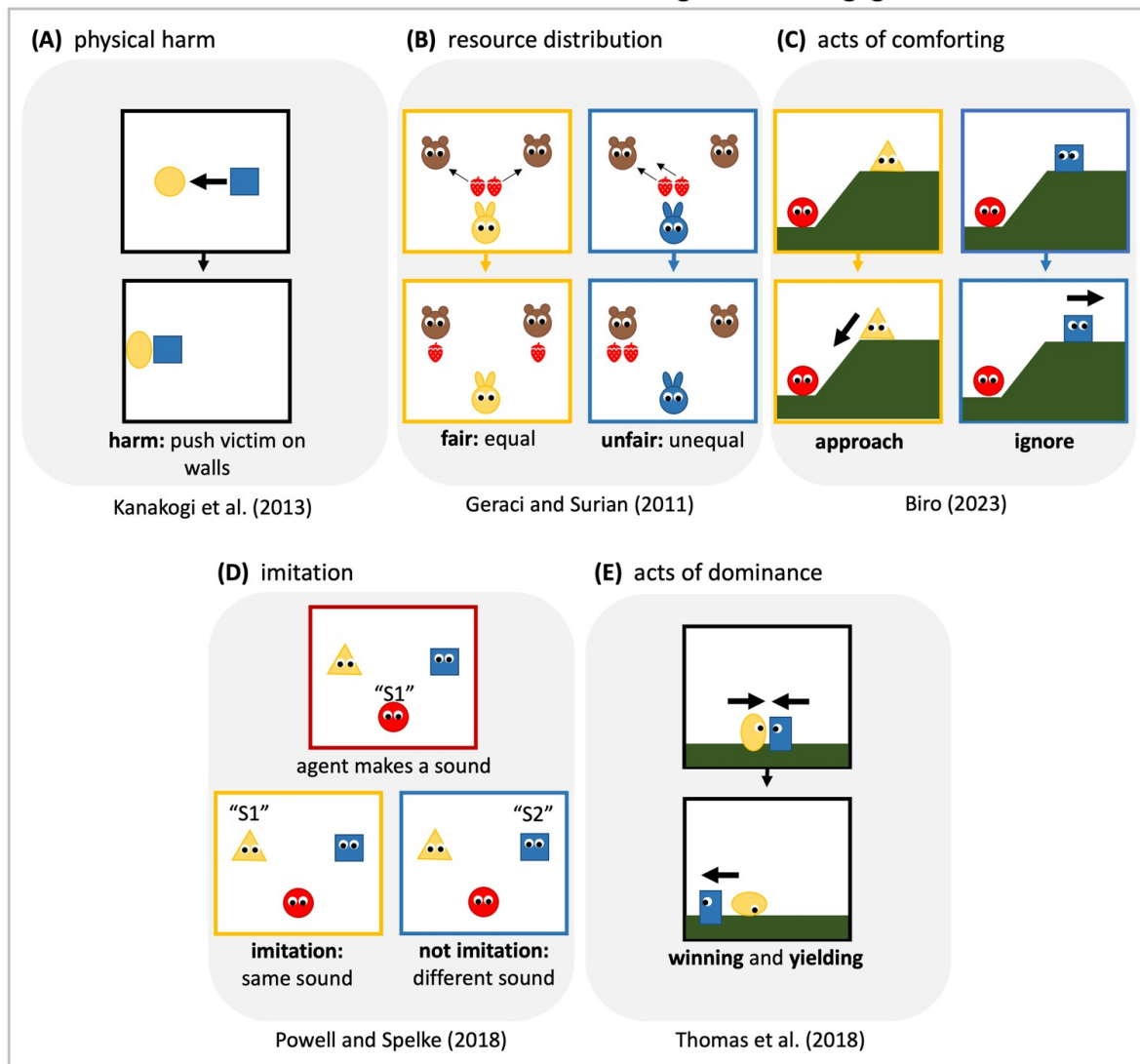


Figure 2. Additional paradigms that have provided evidence that infants and toddlers evaluate social behavior. Specifically, infants and/or toddlers evaluate agents who (A) cause physical harm, (B) distribute resources fairly or unfairly, (C) comfort and ignore others, (D) imitate or do

not imitate others, and (D) win or yield in conflicts. Figures 2A and 2B were adapted with permission from Woo et al. (2022).

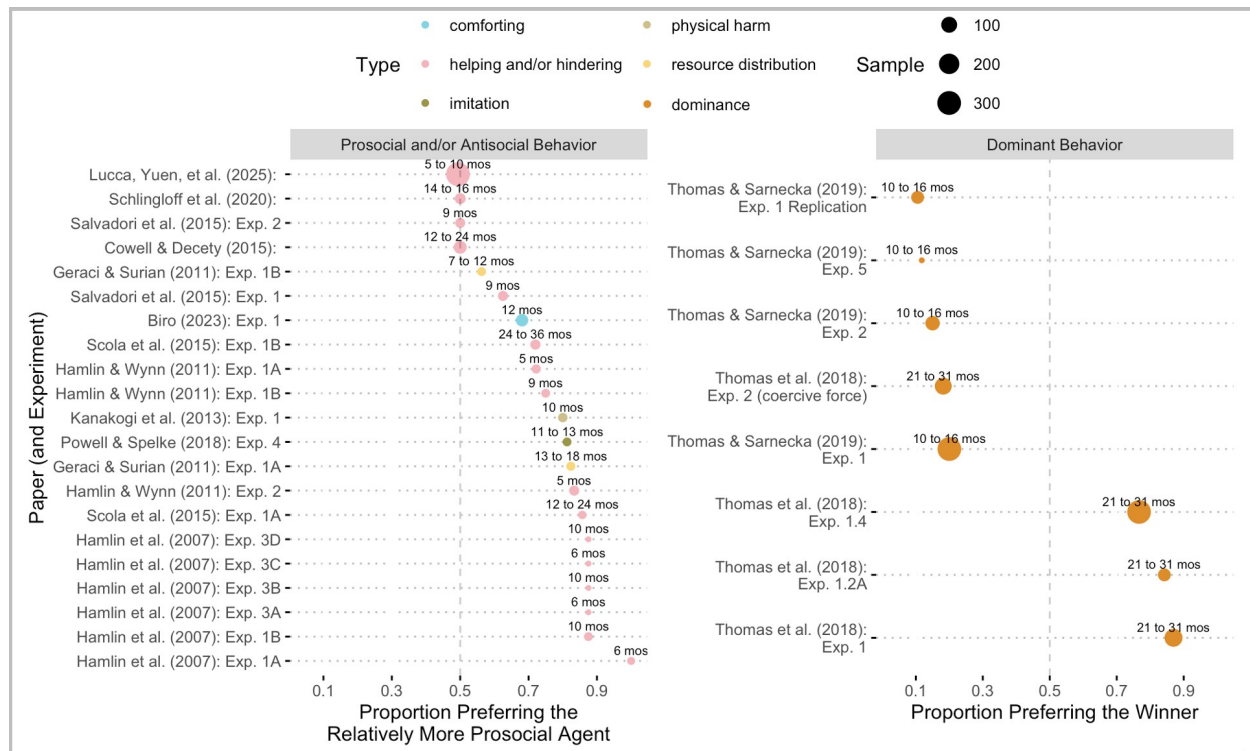


Figure 3. The results of the key papers reviewed concerning infants' and toddlers' evaluations of social agents, as measured by preferential reaching. The size of the circles indicates sample size, color indicates the type of social behavior that a paper involved, and the dashed line indicates chance responding (50%). The age tested in each experiment is written above each circle. The left panel depicts evaluations of behaviors that are prosocial or antisocial, whereas the right panel depicts evaluations of behaviors that involve dominance. We have separated the studies involving dominance from the others, given the dominance studies are not focused on the contrast between prosocial and antisocial agents. We include in this figure the key papers that have introduced paradigms depicting prosocial and antisocial behavior or dominant behavior, as well as attempts

to replicate those findings from independent groups using preferential reaching measures; we found these attempts from examining the papers that have cited the original papers. The data used to generate this graph are hosted on the Open Science Framework: https://osf.io/mhtu8/?view_only=10452caf263b4d2aad37bba090d2ae74.

Acts of Physical Harm. In one experiment, Kanakogi et al. (2013) presented 10-month-old infants with one agent who repeatedly crashed into the second agent, as though the first agent (an aggressor) was hitting or harming the second agent (a victim) (Fig. 2A). After viewing these events, infants preferentially reached for the victim (see Fig. 3). In a further experiment, infants preferentially reached for the victim over a neutral character, and a neutral character over the aggressor. Complementary findings in infants and/or toddlers have come from the same research group (Kanakogi et al., 2017, 2022), as well as by at least three other groups (Buon et al., 2014; Geraci et al., 2024; Thomas et al., 2018). Together, these findings suggest that human infants and toddlers evaluate acts of harm.

Acts of Resource Distribution. Geraci and Surian (2011) presented 7- to 18-month-old infants and toddlers with displays in which one agent distributed two resources equally between two recipients, and another agent instead distributed the resources unequally (Fig. 2B). After viewing these events, the older infants and toddlers (from 13 to 18 months of age) preferentially reached for the agent who distributed resources equally (i.e., a fair distributor) over the agent who distributed resources unequally (i.e., an unfair distributor; see Fig. 3). Although Geraci and Surian found that the younger infants (from 7 to 12 months of age) did not demonstrate a preference, Geraci and colleagues have since reported evidence that 4- and 9-month-old infants evaluate acts of resource distribution (Geraci et al., 2022; Geraci & Surian, 2023). Thus, there is

a discrepancy between the earlier findings of Geraci and Surian (2011) and their later findings. This discrepancy may reflect that early social evaluation is fragile or unreliable, that there may be sampling error (Margoni & Shepperd, 2020), or that there may be methodological differences (see additional discussion below).

We note that these findings largely come from the work of Geraci, of Surian, and their colleagues, but that there has since been evidence from other groups that toddlers evaluate agents who distribute resources equally vs. unequally; specifically, there is evidence that toddlers prefer reaching for toys from a fair distributor over an unfair distributor (Burns & Sommerville, 2014). In a meta-analysis of all the preferential reaching data at the time, Margoni and Surian (2018) found that infants and toddlers prefer fair over unfair distributors. We note that seven years have passed following this meta-analysis. Still, the full set of findings suggests that infants and toddlers evaluate fair and unfair behavior.

Comforting. There is one paper that suggests that infants also evaluate agents who comfort others. Biro (2023) presented 12-month-old infants with one agent who approached a protagonist who was crying (i.e., comforting) and another agent who instead moved away from a protagonist who was crying (i.e., ignoring) (Fig. 2C). Here, infants preferentially reached for the comforter over the ignorer (see Fig. 3). These findings are consistent with the possibility that infants evaluate comforting (Thomas et al., 2020). However, there is an alternative explanation: Infants may have been simply evaluating the agents' approach vs. ignorance behavior (for evidence that newborns preferentially attend to approach behavior, see Geraci et al., 2025). Regardless of the basis of infants' evaluations, they appear to engage in social evaluation. Still,

we look forward to research that continues testing infants' and toddlers' evaluations of comforting and of more general approach vs. ignorance behavior.

Additional Social Behaviors. Finally, beyond behaviors that appear to adults to be prosocial or antisocial, there are other social behaviors that infants and toddlers evaluate. These behaviors include acts of imitation and dominant and submissive behavior.

Imitation. In a series of experiments, Powell and Spelke (2018) presented infants with agents who either imitated others or did not imitate others, or who were either imitated by or not imitated by others. For example, in one experiment ($n = 48$), 4- to 5-month-old infants saw a protagonist produce one sound, one agent (an imitator) subsequently producing the same sound, and another agent (a non-imitator) subsequently producing a different sound (Fig. 2D). In this experiment, as well as an internal replication ($n = 24$), infants preferentially looked to the imitator over the non-imitator. Additionally, 11- to 13-month-old infants preferentially reached for an agent who imitated another's motion over an agent who did not (see Fig. 3). However, in another experiment ($n = 72$), 4- to 5-month-old infants did not prefer an agent who had been imitated over an agent who had not been imitated (cf. Thomas et al., 2022). Thus, infants appear to evaluate agents who do and do not imitate others (for conceptual replication in 12-month-old infants, see Thomas et al., 2020).

Dominant and Submissive Behavior. In one set of experiments, Thomas et al. (2018) presented 21- to 31-month-old toddlers with two agents who were in each other's way, heading in opposite directions. In one experiment, one agent yielded of its own accord, allowing the other agent to continue (Fig. 2E); here, toddlers preferred the agent who "won" in the conflict over the agent who yielded (the "loser"; see Fig. 3). In two further experiments, the researchers internally

replicated this finding. However, when the winner got its way by force, knocking the loser over, toddlers instead preferred the loser over the winner; these findings may reflect that toddlers negatively evaluate agents who engage in physical harm, as reviewed above. Strikingly, when 10- to 16-month-old infants observed the same displays in which one agent got the right of way when another one yielded, as well as displays in which two agents engaged in a dispute over a single toy, infants preferred the agent who yielded (see Fig. 3; Thomas & Sarnecka, 2019). Although it remains unclear as to why infants and toddlers may evaluate the agents differently, these findings suggest that infants and toddlers evaluate dominant and submissive behavior.

Interim Summary. The evidence for early social evaluation does not come only from the one paradigm that the ManyBabies Consortium failed to replicate. There are a number of paradigms, from multiple research groups, that have demonstrated that infants and toddlers engage in social evaluation. We note, however, that for each of these behaviors, they have only been studied by a few particular lab groups (or even just one lab group), and that they have often not been studied by independent research groups. A key direction for future research will be to attempt to build on these findings, through independent groups.

Many Ways to Measure Social Evaluation

To date, the failures to replicate findings that infants and toddlers prefer helpers over hinderers (Lucca, Yuen, et al., 2025; Salvadori et al., 2015; Schlingloff et al., 2020; see also, Margoni & Surian, 2018) have only involved one measure of social evaluation: preferential reaching (Fig. 4A). There are multiple other ways in which researchers have measured infants' and toddlers' social evaluations. The structure of these studies is similar to those of preferential reaching studies, in that infants and toddlers first observe familiarization or habituation events in

which agents engage in social behaviors. The key difference is in the measure of social evaluation.

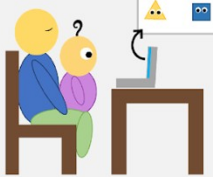
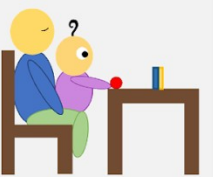
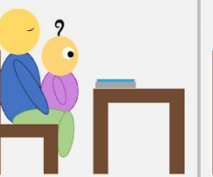
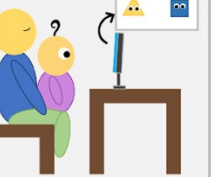
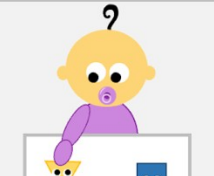
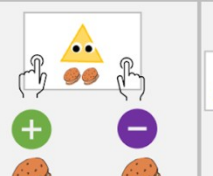
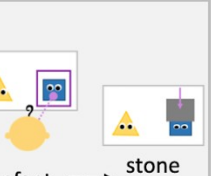
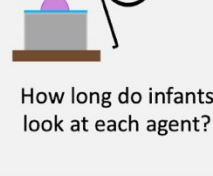
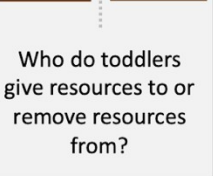
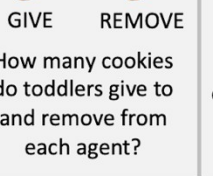
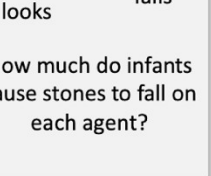
PREFERENCE MEASURES		REWARD AND PUNISHMENT				
		(C) RESOURCE GIVING OR REMOVAL	(D) TABLET	(E) EYETRACKING		
SETUP	(A) PREFERENTIAL REACHING					
	(B) PREFERENTIAL LOOKING					
DEPENDENT VARIABLE	(A) PREFERENTIAL REACHING					
	(B) PREFERENTIAL LOOKING					

Figure 4. The setups and dependent variables for various ways to measure early social evaluation.

Preferential Looking. Preferential looking is perhaps the second most widely used measure of social evaluation, after preferential reaching. After infants and toddlers have observed agents engaged in social actions, rather than presenting participants with two agents to reach for, experimenters present participants with two agents and record participants' looking behavior (Fig. 4B). The experimenters assume that participants' looking behavior reflects their evaluations.

Researchers have used this measure to overcome one of the limitations of preferential reaching as a measure: that preferential reaching requires that an infant or toddler be able to reach for an object. Thus, preferential reaching has not been an appropriate measure of social evaluation for younger infants, who are unable to proficiently reach for and grasp objects. By using preferential looking, studies have been able to examine whether younger infants (e.g., 3- to 5-month-old infants) may look to the same agents that older infants and toddlers reach for. These studies have found that younger infants prefer looking to (i) agents who help over those who hinder (Hamlin et al., 2010; Hamlin & Wynn, 2011), (ii) agents who imitate over those who do not (Powell & Spelke, 2018), and (iii) agents who attempt to distribute resources fairly over agents who attempt to distribute resources unfairly (Geraci et al., 2022).

Whereas preferential reaching has required that infants and toddlers be physically present in the lab to reach for agents or puppets, preferential looking requires no physical contact. When labs closed on-site operations during the COVID-19 pandemic, researchers attempted to measure infants' and toddlers' evaluations through online testing, with preferential looking as a measure. These efforts have provided findings that infants and toddlers prefer looking to prosocial over antisocial agents (Woo et al., 2024; Woo & Spelke, 2023b, 2023a). There is evidence from one paper that, after observing the same stimuli, effect sizes are similar for preferential looking and reaching (Chuey et al., 2021, see also, Chuey et al., 2024).

We note, however, that in the same study ($n = 20$) in which 10-month-old infants preferentially reached for a victim over an aggressor, the infants did not preferentially look at the victim over the aggressor (Kanakogi et al., 2013). Thus, there is at least one datapoint suggestive that preferential looking may not always replicate the findings of preferential reaching.

Reward and Punishment. A growing number of studies have presented infants and toddlers with agents who engage in social behaviors and then examined infants' and toddlers' decisions to reward and punish those agents. Would infants and toddlers preferentially reward prosocial agents, and preferentially punish antisocial agents?

Hamlin et al. (2011) investigated this question in an experiment ($n = 16$ per condition), examining how 19-month-old toddlers would distribute treats or remove treats from agents who had helped or hindered others (Fig. 4C). They found that after viewing helping and hindering events, toddlers preferentially gave treats to the helper (81% of toddlers did so), and preferentially removed treats from the hinderer to give to someone else (87% of toddlers did so; for complementary findings in 20-month-old toddlers, see Van de Vondervoort et al., 2018). These findings suggest that toddlers reward helpers and punish hinderers.

Researchers have since developed additional ways to probe infants' and toddlers' choices to reward and punish others. Ziv et al. (2021) developed a tablet-based task, in which 16-month-old toddlers learned that touches on one side of a screen rewarded an agent (verbal praising the agent or giving the agent a cookie) and that touches on the other side produced punishment (admonishing the agent or removing a cookie from the agent) (Fig. 4D). Building on findings that infants and toddlers evaluate agents who distribute resources fairly vs. unfairly, the experimenters asked whether toddlers would differently reward and punish agents depending on whether those agents had distributed resources fairly vs. unfairly. They found that toddlers preferentially rewarded the fair distributor: The toddlers were more likely (i) to touch the tablet to praise the fair distributor than to admonish the fair distributor in one experiment ($n = 64$) and (ii) to touch the tablet to give the fair distributor a cookie than to take away a cookie from the fair

distributor in a second experiment ($n = 32$). Toddlers did not, however, differently touch the tablet to reward or punish the unfair distributor. These findings suggest that toddlers reward fair distributors, but do not punish unfair distributors.

The earlier reward and punishment studies with toddlers (e.g., Hamlin et al., 2011; Ziv et al., 2021) have depended on toddlers being able to move resources or interface with tablets. The methods within these studies have therefore been unsuitable for younger infants. There have been studies of infants' evaluations of agents who reward and punish helpers and hinderers (e.g., Hamlin et al., 2011) and of infants' expectations for whether individuals will reward or punish fair and unfair distributors (Geraci & Surian, 2023; Meristo & Surian, 2013). But would infants themselves preferentially reward prosocial agents or punish antisocial agents?

Kanakogi et al. (2022) developed a paradigm involving eye tracking, suitable for examining infants' punishment behavior. Here, 8-month-old infants first participated in practice and pretest trials, during which, when they looked at an object or an agent for 500 ms, a stone would fall and crush the object or agent (Fig. 4E). Next, in one experiment ($n = 24$), infants were familiarized to events in which one agent hit or harmed another; that is, one agent was an aggressor and one was a victim (as in Kanakogi et al., 2013). Finally, infants participated in trials in which they viewed the aggressor and the victim, and their looking again caused stones to fall and crush an agent. Kanakogi et al. found that infants were more likely to cause stones to fall on the aggressor. Importantly, in another experiment ($n = 24$), when infants' looking caused a soft object to fall, which did not crush an agent, infants did not look differently to the aggressor and victim. Thus, these findings suggest that infants punish aggressors over victims. Such methods could be adapted to examine whether infants may also reward prosocial over antisocial agents.

Limitations of Preferential Reaching. Preferential reaching has at least four limitations to it. First, as mentioned above, it requires that participants be able to reach, restricting its use to infants who are old enough to manipulate objects and to participants who can be present in person. In contrast, preferential looking has been used on infants who lack the motor abilities to effectively reach for objects, as well as on online samples. Reward and punishment studies based on eye tracking do not depend on reaching, but their use is very novel; future research should adapt such methods for younger infants and for online testing.

Second, preferential reaching is a single binary measure. While a participant's choice to reach for one agent over another may indicate that the participant prefers one agent, it does not shed light on how much a single participant prefers one agent over another. In contrast, preferential looking is a metric measure, from which researchers can gain insight as to how much a participant prefers looking at one agent over another. Likewise, reward and punishing behavior can provide metric data: how much a participant chooses to reward an agent (e.g., the number of cookies distributed) or punish an agent (e.g., the number of times admonishment is played or a stone falls on the agent). Relative to a single binary measure, repeated looking, repeated reward behavior, and repeated punishment behavior may be more sensitive measures.

Third, preferential reaching studies may be especially dependent on participants being comfortable. Although all of the kinds of studies that we have reviewed are dependent on participants being comfortable (i.e., not fussing out of a study), preferential reaching has an additional feature: For infants or toddlers to preferentially reach for one of two agents, an experimenter must typically present those agents. Infants at these ages can reach for targets and may see and interact with adults aside from their own caregivers; however, when they do not

know or are uncomfortable with an experimenter, infants and toddlers may be less likely to reach for an agent. The ages tested in studies of social evaluation often coincide with ages at which infants experience wariness of strangers (Brooker et al., 2013; Sroufe, 1977). Thus, for preferential reaching studies, there are often warm-up periods, before an experiment, to introduce infants and toddlers to the experimenter. But these warm-up periods could look very different from one lab to the next, or even from one infant to the next in the same lab. That is, preferential reaching studies may be highly variable between participants in terms of participants' level of engagement with the experimenter. Some research has revealed that warm-up activities can influence the rate at which toddlers help others (Cortes Barragan & Dweck, 2014); it is unknown whether warm-up activities may influence infants' or toddlers' reaching behavior. In contrast to preferential reaching studies, preferential looking studies and reward and punishment studies conducted over screens or eye tracking do not require participants to interact as closely with an experimenter.

Finally, preferential reaching studies sometimes present participants with 2D videos of characters in familiarization, followed by the 3D versions of those characters that the participants can reach for. For example, in Lucca, Yuen, et al. (2025), infants observed videos depicting wooden characters involved in helping and hindering actions, and then reached for 3D foam versions of those characters. The shift from 2D to 3D, using different materials, may present a challenge for participants: They need to infer that the later physical versions of the characters are the same characters that they saw in the 2D videos. This may be a challenge: There is some evidence that 19-month-old toddlers do not connect events that they see on 2D screens to events in real life (Preissler & Carey, 2004; Revencu & Csibra, 2021). Moreover, participants may be challenged to track multiple agents or objects (see Carey & Xu, 2001; Xu & Carey, 1996).

Preferential looking studies and reward and punishment studies may avoid these two challenges, when agents are presented in the same way (e.g., always 2D on a screen) during familiarization videos that depict social actions and during the later tasks.

Preferential looking and reward and punishment studies offer multiple advantages over preferential reaching. If preferential reaching is not a replicable measure of social evaluation, then preferential looking and reward and punishment studies may be appropriate alternatives for developmental psychologists moving forward.

Looking Forward

The original finding that infants prefer reaching to helpers over hinderers has not been replicated in multiple studies, , and these failed replications suggest that early social evaluation may be fragile or unreliable, if at all present. In the present paper, we have noted that the bulk of these null findings have come from just one paradigm (involving a protagonist who tries to climb a hill), and from using just one dependent variable (preferential reaching). There is evidence that infants and toddlers evaluate a wider range of social behaviors, and there are multiple other ways to study social evaluation, such as preferential looking and reward and punishment behaviors. The larger body of evidence therefore suggests that early social evaluation is a real phenomenon.

We see four key areas for future research. First, although there is a large body of evidence that infants and toddlers evaluate a wide range of social behaviors, within each individual paradigm (e.g., evaluations of fair and unfair behavior), perhaps only a few groups have studied each kind of behavior. We look forward to research that explores the replicability of these other paradigms, beyond the original helping and hindering work of Hamlin et al. (2007). Such independent replication attempts will be crucial tests of the possibility that infants and toddlers engage in social evaluation.

Second, although infants and toddlers often appear to preferentially look at more prosocial agents, our field should consider conducting a meta-analysis of preferential looking findings and examining the replicability of these findings. The same applies to reward and punishment studies, perhaps even more so, given that a relatively smaller number of studies have used such methods, with the scope of the behaviors presented to infants and toddlers being similarly limited.

Third, although researchers have collectively examined infants' and toddlers' evaluations of a wide range of behaviors, it is sometimes unclear whether infants' and toddlers' responses across these studies reflect the development of knowledge in the same system or in different systems. That is, as adults, we could describe helping, being fair, and protecting others as being prosocial. But the infants and toddlers who have participated in studies examining early evaluations of these behaviors have almost always been different individual participants. Although adults see connections between these behaviors, do infants and toddlers? Is there some abstract "prosocial" quality linking these behaviors to each other? Studies are beginning to examine interrelations between these different behaviors (Gill & Sommerville, 2023; Kudrnova et al., 2024; Pepe & Powell, 2025; Surian et al., 2018). In one such study, for example, Gill and Sommerville presented 14- to 27-month-old infants and toddlers with an agent who distributed resources unfairly. In test trials, the unfair distributor either helped or hindered a protagonist in the pursuit of a goal. Here, the infants and toddlers looked longer when the unfair distributor helped than when the unfair distributor hindered. These findings suggest that from an agent's antisocial behavior (distributing resources unfairly), infants and toddlers form inferences that the agent will behave antisocially in other ways (e.g., hindering). Future research could examine whether the same infants who evaluate one social behavior (e.g., preferring a fair distributor over

an unfair one) may also be more likely to evaluate other social behaviors (e.g., preferring imitators over non-imitators).

Finally, although nonverbal behaviors (e.g., reaching, looking, etc.) can shed light on infants' and toddlers' developing knowledge, these measures are still ambiguous. For example, although touching someone involves approaching one agent while avoiding another and could therefore reflect a social preference, reaching may also be a precursor to behaviors that suggest that someone dislikes another, such as pushing away or hitting. Likewise, looking can reflect not just preference or interest, but also fear and wariness (Rakison & Derringer, 2008). Future research should disambiguate the reasons that infants and toddlers reach for and look at agents.

One possibility is that infants reach for and look at prosocial agents because they see those agents as likely to behave prosocially towards others, including new social partners. Behavioral research is already testing this possibility by examining infants' and toddlers' expectations of agents who engage in prosocial and antisocial actions: how they expect these agents to later interact with each other and with novel social partners (Gill & Sommerville, 2023; Jin et al., 2024; Kudrnova et al., 2024; Pepe et al., 2025; Pepe & Powell, 2025; Surian et al., 2018).

Looking forward, neuroimaging research and longitudinal research could continue to investigate the reasons why infants may engage in preferential reaching and looking. One open question is whether the same regions of the brain that support adults' social reasoning are also engaged when infants observe prosocial and antisocial behavior (Farris et al., 2022; Gredebäck et al., 2015; Hyde et al., 2018; Tan & Hamlin, 2022). Such investigations could shed light on whether infants' behavioral responses to social actions reflects social reasoning, as well as how

the infant social mind compares to that of adults. Researchers may consider connecting infants' neural responses to their behavior.

Regarding longitudinal research, it remains largely unknown whether and how much infants' and toddlers' behavioral responses to social interactions predict later social and moral development (for evidence regarding continuity in physical reasoning, see Perez & Feigenson, 2021). Such research could shed light on continuity in social reasoning. There is some evidence linking early evaluations of helping and hindering with later social functioning (Tan et al., 2018), but less is known concerning whether and how much infants' and toddlers' responses to other social behaviors predict later development.

Conclusion

The study of early social evaluation has experienced a replication crisis, and this replication crisis has inspired questions whether early social evaluation is real. Here, we have sought to contextualize this replication crisis: that this replication crisis is relatively contained, focused on a single paradigm (involving a protagonist who is climbing a hill) and a single dependent variable (preferential reaching). Beyond this single paradigm, there are multiple other paradigms from which there is evidence that infants and toddlers engage in social evaluation. And beyond preferential reaching, infants and toddlers engage in preferential looking and reward and punishment behavior. Following the present replication crisis, we should not abandon the study of early social evaluation. Instead, we should treat this as an opportunity to reevaluate how we have studied early social evaluation, so that we can improve our methods for future research. This future research will continue to characterize what infants and toddlers may already understand of the social world, and this understanding may serve as the foundation of humans' social minds.

References

- Biro, S. (2023). Twelve months old infants' evaluation of observed comforting behavior using a choice paradigm: The role of animacy cues and self-distress. *Infancy*, 28(4), 807–819.
<https://doi.org/10.1111/infa.12542>
- Brooker, R. J., Buss, K. A., Lemery-Chalfant, K., Aksan, N., Davidson, R. J., & Goldsmith, H. H. (2013). The development of stranger fear in infancy and toddlerhood: Normative development, individual differences, antecedents, and outcomes. *Developmental Science*, 16(6), 864–878. <https://doi.org/10.1111/desc.12058>
- 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), Article 2.
- Burns, M. P., & Sommerville, J. (2014). “I pick you”: The impact of fairness and race on infants' selection of social partners. *Frontiers in Psychology*, 5, 93.
- Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, 80(1–2), 179–213.
- Chan, A. K. (2002). *Mencius: Contexts and interpretations*. University of Hawaii Press.
https://books.google.com/books?hl=en&lr=&id=QviMzbBz7e0C&oi=fnd&pg=PP13&dq=chan+2002+mencius&ots=Eprvb3_1_a&sig=vwCjVNaR4owuA9eVYiX9krU8M2M
- Chuey, A., Asaba, M., Bridgers, S., Carrillo, B., Dietz, G., Garcia, T., Leonard, J. A., Liu, S., Merrick, M., & Radwan, S. (2021). Moderated online data-collection for developmental research: Methods and replications. *Frontiers in Psychology*, 4968.
- Chuey, A., Boyce, V., Cao, A., & Frank, M. C. (2024). Conducting developmental research online vs. in-person: A meta-analysis. *Open Mind*, 8, 795–808.

- Cortes Barragan, R., & Dweck, C. S. (2014). Rethinking natural altruism: Simple reciprocal interactions trigger children's benevolence. *Proceedings of the National Academy of Sciences*, 111(48), 17071–17074. <https://doi.org/10.1073/pnas.1419408111>
- Cowell, J. M., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *Proceedings of the National Academy of Sciences*, 112(41), Article 41. <https://doi.org/10.1073/pnas.1508832112>
- Descartes, R. (2012). *Meditation on first philosophy*. Simon and Schuster.
<https://books.google.com/books?hl=en&lr=&id=KlTrAgAAQBAJ&oi=fnd&pg=PT9&dq=Meditations+on+First+Philosophy&ots=qfrAA7YECk&sig=vQda7pa9VfHK7tTUxOZ2mtP2r6g>
- Dewey, J. (2002). *Human nature and conduct*. Courier Corporation.
<https://books.google.com/books?hl=en&lr=&id=2M47lIDaQwQC&oi=fnd&pg=PA1&dq=Human+nature+and+conduct&ots=qFGuZmcmIy&sig=zTOTUb0atDzoutdu4O3kYiiMz4w>
- Farris, K., Kelsey, C. M., Krol, K. M., Thiele, M., Hepach, R., Haun, D. B., & Grossmann, T. (2022). Processing third-party social interactions in the human infant brain. *Infant Behavior and Development*, 68, 101727.
- Geraci, A., Benavides-Varela, S., Nascimben, C., Simion, F., & Di Giorgio, E. (2024). Evaluations of aggressive chasing interactions by 7-month-old infants. *Aggressive Behavior*, 50(5), e22174. <https://doi.org/10.1002/ab.22174>
- Geraci, A., Simion, F., & Surian, L. (2022). Infants' intention-based evaluations of distributive actions. *Journal of Experimental Child Psychology*, 220, 105429.
<https://doi.org/10.1016/j.jecp.2022.105429>

- Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, 14(5), Article 5.
- Geraci, A., & Surian, L. (2023). Intention-based evaluations of distributive actions by 4-month-olds. *Infant Behavior and Development*, 70, 101797.
- Geraci, A., Surian, L., Tina, L. G., & Hamlin, J. K. (2025). Human newborns spontaneously attend to prosocial interactions. *Nature Communications*, 16(1), 1–9.
<https://doi.org/10.1038/s41467-025-61517-3>
- Gill, I. K., & Sommerville, J. A. (2023). Generalizing across moral sub-domains: Infants bidirectionally link fairness and unfairness to helping and hindering. *Frontiers in Psychology*, 14, 1213409.
- Gredebäck, G., Kaduk, K., Bakker, M., Gottwald, J., Ekberg, T., Elsner, C., Reid, V., & Kenward, B. (2015). The neuropsychology of infants' pro-social preferences. *Developmental Cognitive Neuroscience*, 12, 106–113. <https://doi.org/10.1016/j.dcn.2015.01.006>
- Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, 22(3), Article 3.
<https://doi.org/10.1177/0963721412470687>
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26(1), Article 1. <https://doi.org/10.1016/j.cogdev.2010.09.001>
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450(7169), Article 7169. <https://doi.org/10.1038/nature06288>
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, 13(6), Article 6.
<https://doi.org/10.1111/j.1467-7687.2010.00951.x>

- 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(50), Article 50.
<https://doi.org/10.1073/pnas.1110306108>
- Hobbes, T. (1967). *Leviathan*. Рипол Классик. https://books.google.com/books?hl=en&lr=&id=oLANAwAAQBAJ&oi=fnd&pg=PA9&dq=leviathan+hobbes&ots=nqQpqUXzQi&sig=S9AODajpcCkXoH_LWiLjfcIR_s
- Hume, D. (2003). *A treatise of human nature*. Courier Corporation.
- Hyde, D. C., Simon, C. E., Ting, F., & Nikolaeva, J. I. (2018). Functional Organization of the Temporal–Parietal Junction for Theory of Mind in Preverbal Infants: A Near-Infrared Spectroscopy Study. *Journal of Neuroscience*, 38(18), Article 18.
<https://doi.org/10.1523/JNEUROSCI.0264-17.2018>
- Jin, K., Ting, F., He, Z., & Baillargeon, R. (2024). Infants expect some degree of positive and negative reciprocity between strangers. *Nature Communications*, 15(1), 7742.
- 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), Article 2.
- Kanakogi, Y., Miyazaki, M., Takahashi, H., Yamamoto, H., Kobayashi, T., & Hiraki, K. (2022). Third-party punishment by preverbal infants. *Nature Human Behaviour*, 6(9), 1234–1242.
<https://doi.org/10.1038/s41562-022-01354-2>
- 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), Article 6.
- Kant, I. (1908). Critique of pure reason. *Modern Classical Philosophers*, Cambridge, MA: Houghton Mifflin, 370–456.

- Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, 12(2), 93–120.
- Kudrnova, V., Spelke, E. S., & Thomas, A. J. (2024). Infants Infer Social Relationships Between Individuals Who Engage in Imitative Social Interactions. *Open Mind*, 8, 202–216.
https://doi.org/10.1162/opmi_a_00124
- Locke, J. (1847). *An essay concerning human understanding*. Kay & Troutman.
- Lucca, K., Yuen, F., Wang, Y., Alessandroni, N., Allison, O., Alvarez, M., Axelsson, E. L., Baumer, J., Baumgartner, H. A., Bertels, J., Bhavsar, M., Byers-Heinlein, K., Capelier-Mourguy, A., Chijiwa, H., Chin, C. S. -S., Christner, N., Cirelli, L. K., Corbit, J., Daum, M. M., ... Hamlin, J. K. (2025). Infants' Social Evaluation of Helpers and Hinderers: A Large-Scale, Multi-Lab, Coordinated Replication Study. *Developmental Science*, 28(1), e13581. <https://doi.org/10.1111/desc.13581>
- Margoni, F., & Shepperd, M. (2020). Changing the logic of replication: A case from infant studies. *Infant Behavior and Development*, 61, 101483.
- Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology*, 54(8), Article 8.
- Melis, A. P., & Raihani, N. J. (2023). The cognitive challenges of cooperation in human and nonhuman animals. *Nature Reviews Psychology*, 2(9), 523–536.
- Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition*, 129(1), 102–113. <https://doi.org/10.1016/j.cognition.2013.06.006>
- Pepe, B., & Powell, L. J. (2025). Infants' Expectations for Helping in Imitators. *Infancy*, 30(1), e12642. <https://doi.org/10.1111/infa.12642>

- Pepe, B., Woo, B. M., Thomas, A. J., & Powell, L. J. (2025). Helping and hindering guide infants' expectations about future behavior. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society*.
- Perez, J., & Feigenson, L. (2021). Stable individual differences in infants' responses to violations of intuitive physics. *Proceedings of the National Academy of Sciences*, 118(27), e2103805118. <https://doi.org/10.1073/pnas.2103805118>
- Piaget, J. (1965). *The moral judgment of the child*. Routledge & K. Paul.
- Powell, L. J., & Spelke, E. S. (2018). Third-party preferences for imitators in preverbal infants. *Open Mind*, 2(2), Article 2.
- Preissler, M. A., & Carey, S. (2004). Do Both Pictures and Words Function as Symbols for 18- and 24-Month-Old Children? *Journal of Cognition and Development*, 5(2), 185–212. https://doi.org/10.1207/s15327647jcd0502_2
- Rakison, D. H., & Derringer, J. (2008). Do infants possess an evolved spider-detection mechanism? *Cognition*, 107(1), 381–393.
- Revenu, B., & Csibra, G. (2021). For 19-month-olds, what happens on-screen stays on-screen. *Open Mind*, 5, 71–90.
- Rottman, J., & Young, L. (2015). Mechanisms of Moral Development. In *The moral brain: A multidisciplinary perspective* (p. 20). MIT Press.
- Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PloS One*, 10(11), Article 11.
- Schlingloff, L., Csibra, G., & Tatone, D. (2020). Do 15-month-old infants prefer helpers? A replication of Hamlin et al.(2007). *Royal Society Open Science*, 7(4), Article 4.

Schmidt, M. F. H., & Rakoczy, H. (2023). Children's Acquisition and Application of Norms.

Annual Review of Developmental Psychology, 5(1), 193–215.

<https://doi.org/10.1146/annurev-devpsych-120621-034731>

Scola, C., Holvoet, C., Arciszewski, T., & Picard, D. (2015). Further Evidence for Infants'

Preference for Prosocial Over Antisocial Behaviors. *Infancy*, 20(6), 684–692.

<https://doi.org/10.1111/infa.12095>

Sommerville, J. A. (2024). The origins of moral sensitivities: Probing infants' expectations,

evaluations, generalization, and enforcement of moral norms. *Adv Child Dev Behav*, 67,

31–69.

Sroufe, L. A. (1977). Wariness of strangers and the study of infant development. *Child*

Development, 731–746.

Steckler, C. M., Woo, B. M., & Hamlin, J. K. (2017). The limits of early social evaluation: 9-

month-olds fail to generate social evaluations of individuals who behave inconsistently.

Cognition, 167, 255–265. <https://doi.org/10.1016/j.cognition.2017.03.018>

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.

Tan, E., & Hamlin, J. K. (2022). Infants' neural responses to helping and hindering scenarios.

Developmental Cognitive Neuroscience.

Tan, E., Mikami, A. Y., & Hamlin, J. K. (2018). Do infant sociomoral evaluation and action

studies predict preschool social and behavioral adjustment? *Journal of Experimental*

Child Psychology, 176, 39–54. <https://doi.org/10.1016/j.jecp.2018.07.003>

- Thomas, A. J., & Sarnecka, B. W. (2019). Infants choose those who defer in conflicts. *Current Biology*, 29(13), Article 13.
- Thomas, A. J., Saxe, R., & Spelke, E. (2020). Infants use imitation but not comforting or social synchrony to evaluate those insocial interactions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 42. <https://escholarship.org/uc/item/1dd1r4gb>
- Thomas, A. J., Thomsen, L., Lukowski, A. F., Abramyan, M., & Sarnecka, B. W. (2018). Toddlers prefer those who win but not when they win by force. *Nature Human Behaviour*, 2(9), Article 9. <https://doi.org/10.1038/s41562-018-0415-3>
- Ting, F., Dawkins, M. B., Stavans, M., & Baillargeon, R. (2019). Principles and concepts in early moral cognition. *The Social Brain: A Developmental Perspective*, 41–65.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Van de Vondervoort, J. W., Akin, L. B., Kushnir, T., Slevinsky, J., & Hamlin, J. K. (2018). Selectivity in toddlers' behavioral and emotional reactions to prosocial and antisocial others. *Developmental Psychology*, 54(1), 1.
- Woo, B. M., Liu, S., Gweon, H., & Spelke, E. S. (2024). Toddlers Prefer Agents Who Help Those Facing Harder Tasks. *Open Mind*, 8, 483–499. https://doi.org/10.1162/opmi_a_00129
- Woo, B. M., & Spelke, E. S. (2023a). Infants and toddlers leverage their understanding of action goals to evaluate agents who help others. *Child Development*, 94(3), 734–751. <https://doi.org/10.1111/cdev.13895>
- Woo, B. M., & Spelke, E. S. (2023b). Toddlers' social evaluations of agents who act on false beliefs. *Developmental Science*, 26(2), e13314. <https://doi.org/10.1111/desc.13314>

- Woo, B. M., Tan, E., & Hamlin, J. K. (2022). Human Morality Is Based on an Early-Emerging Moral Core. *Annual Review of Developmental Psychology*, 4(1), 41–61.
<https://doi.org/10.1146/annurev-devpsych-121020-023312>
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, 30(2), 111–153.
- Ziv, T., Whiteman, J. D., & Sommerville, J. A. (2021). Toddlers' interventions toward fair and unfair individuals. *Cognition*, 214, 104781.