Title

Eleven-Month-Old Infants Preferentially Look at Helpers But Do Not Reliably Evaluate Inconsistent Actors

Journal

Open Mind, 10

ISSN

2470-2986

Authors

Woo, Brandon M

Tsang, Adrian

Hamlin, J Kiley

Publication Date

2026-07-17

DOI

10.1162/opmi.a.370

Eleven-Month-Old Infants Preferentially Look at Helpers But Do Not Reliably Incorporate Inconsistency Into Their Evaluations

Brandon M. Woo[1], Adrian Tsang[2], and J. Kiley Hamlin[2]

[1]Department of Psychological and Brain Sciences, University of California, Santa Barbara

[2]Department of Psychology, University of British Columbia

Abstract

Adults quickly form character impressions from their observations of others' behavior and sometimes manage to do so even when an agent's behaviors are inconsistent. Developmental psychology has provided some evidence that infants, like adults, may form impressions based on prosocial and antisocial acts: They prefer those who behave prosocially versus antisocially toward third parties. That said, previous work suggests that infants find it difficult to evaluate characters' whose behaviors are inconsistent. The present, preregistered experiments further explored this phenomenon in an older age group (11-month-olds; N = 189) using an online format and presenting infants with giving versus taking, a different type of prosocial versus antisocial action than in previous work. In a preliminary experiment, we successfully replicated the finding that infants prefer looking to helpful givers over unhelpful takers when actors behave consistently and when each actor performed 4 actions. We then used the same online testing methods to examine infants' evaluations of inconsistent actors. In Experiment 1, infants appeared able to evaluate inconsistent actors, preferring more versus less prosocial actors when their inconsistent act occurred last in a series of 5 acts; however, they did not appear to find the behavioral inconsistency unexpected, suggesting they may have failed to notice it. In Experiment 2, we adjusted the order of inconsistent acts and put the inconsistent act first, and attempted to conceptually replicate the findings of Experiment 1. Here, infants did not prefer more prosocial actors. Together with past work, these findings suggest that while infants prefer agents whose actions are consistently prosocial versus antisocial, they do not reliably incorporate behavioral inconsistency into their social evaluations.

Key words: cognitive development; social evaluation; impression formation; inconsistent behavior

Eleven-Month-Old Infants Preferentially Look at Helpers But Do Not Reliably Incorporate Inconsistency Into Their Evaluations

For years, Ellen Degeneres regularly highlighted the work of local community heroes on her talk show. Through these and other actions, she built a reputation on being a kind and generous celebrity. Everything changed, however, when the news broke that Degeneres had allowed, or even created, a culture of bullying and harassment among the employees of her show (e.g., firing employees who took medical leave). Many people updated their impressions of Degeneres, and her long-running talk show ended. In everyday life, a given individual (like Degeneres) may act both prosocially and antisocially over time. To generate accurate impressions, then, one must aggregate across different and potentially conflicting behaviors in order to form holistic evaluations of an actor's character. The present paper investigates the development of holistic impression formation, by assessing whether and how infants evaluate actors who behave inconsistently.

Impression Formation in Adults

By observing other people's behavior, adults quickly form character impressions. Research suggests this process is both persistent and intuitive: Adults generate impressions even from sparse behavioral data (Ambady et al., 1995; Ambady & Rosenthal, 1992), even when they are not trying to (Uleman et al., 2005; Winter & Uleman, 1984), and even when there are alternative, external explanations for an individual's actions (Brosch et al., 2013; Gilbert & Malone, 1995; Jones & Harris, 1967; Kubota et al., 2014; Ross, 1977).

These persistent character impressions may be adaptive. Indeed, generating theories of others' character from overt behaviors allows social agents to adjust their own social decision-making accordingly; for instance, deciding to interact with those who behave prosocially while

avoiding those who have behaved antisocially. These "partner choices" ultimately support human cooperation, by allowing humans to selectively cooperate with those likely to cooperate in return (Cosmides & Tooby, 1992; Darwin, 2008; Henrich & Henrich, 2007; Joyce, 2007; Katz, 2000). But of course, people are not always consistent in their behavior, rendering adaptive partner choice more complicated. How do humans make sense of such inconsistency?

Perhaps an optimal strategy for impression formation would be to maintain a running average of each individual's positive and negative behaviors. From this average, one could easily determine an individual's current social value by positively evaluating those with net-positive averages and negatively evaluating those with net-negative ones. Despite the intuitive (and seemingly useful) nature of this averaging process, decades of research in social psychology have demonstrated that adults simply fail to use it, and instead have various (and sometimes incompatible) impression-formation biases that resist summarizing and/or updating. For example, there is evidence that adults may rely on first impressions (i.e., a primacy bias; Anderson, 1965; Anderson & Hubert, 1963; see also, Kim et al., 2020, 2022), and that adults may place more weight on negative behaviors over positive ones (Baumeister et al., 2001; Ferguson et al., 2019; Kanouse & Hanson Jr., 1987; Rozin & Royzman, 2001; Skowronski & Carlston, 1989).

How and why do capacities for impression formation, alongside difficulty with or resistance to engaging in basic averaging processes, emerge in human development? At least two, non-mutually exclusive possibilities exist. One possibility is that adults engage in rapid, persistent, and sometimes biased impression formation due to a lifetime of experience observing and making sense of others' positive and negative behaviors. Alternatively, capacities for impression formation may be based on early-emerging theories about the social world, which are

then reinforced and/or shifted by experience. If so, it is possible that such capacities may emerge independently of experience.

Impression Formation in Infancy

Do preverbal infants, with relatively limited experience with others' behaviors – and particularly with how individuals' behaviors vary across time – nevertheless use observed behavior to generate impressions? A growing literature suggests that infants may use others' consistent social behaviors to inform their partner choices, preferring to look at and reach for individuals who previously helped versus hindered another's goals within the first 3-6 months after birth (Hamlin et al., 2007, 2010; Hamlin & Wynn, 2011). In some cases, infants have demonstrated reliable prosocial preferences based on as little as a single act of helping or hindering (Hamlin, 2014; Hamlin et al., 2011). Although this body of research includes several failed replications (Cowell & Decety, 2015; Lucca, Yuen, et al., 2025; Nighbor et al., 2017; Salvadori et al., 2015; Schlingloff et al., 2020), there have also been numerous replications and extensions (for meta-analyses, see Margoni & Surian, 2018; Tan, 2024). Indeed, research has additionally revealed that infants' and toddlers' evaluate individuals who behave fairly versus unfairly (Burns & Sommerville, 2014; Geraci, 2022; Geraci et al., 2022; Geraci & Surian, 2011, 2023; Lucca et al., 2018) and who engage in physical harm and protection (Buon et al., 2014; Kanakogi et al., 2013, 2017). These findings suggest (i) that infants' prosocial preferences apply to numerous forms of prosocial and antisocial behaviors and (ii) that there may be early-emerging abilities to evaluate individuals based on their prosocial and antisocial acts (for review, see Hamlin, 2013; Woo et al., 2022). These early-emerging evaluations may be a precursor to adults' capacities for impression formation.

But do these evaluations reflect that infants are forming *impressions* of others based on their behaviors? If infants are forming impressions, then they should expect others to behave consistently over time. Indeed, a number of studies suggest that both infants and toddlers expect consistency in prosocial and antisocial behavior. One group of studies exploring this question has provided evidence that infants expect agents to be consistent in their behaviors *toward the same target*. For instance, Shimizu and colleagues (2018) found that both infants (6-, 9-, and 12-month-olds) and toddlers (15- to 18-month-olds) expected consistency in valenced social behaviors, looking longer when an actor who had previously hindered a target (e.g., prevented them from opening a box) later helped them (e.g., returned their dropped ball). That is, infants and toddlers appeared to expect actors to treat a given target consistently over time, across the same as well as distinct instances of pro- and anti-social behaviors (for complementary findings, see Taborda-Osorio et al., 2019; Tatone et al., 2015; see also, Pepe et al., 2025). There is also some evidence that by some time in the second year of life, infants may use an actor's prosocial and/or antisocial behaviors toward one target to generate expectations for how they will behave toward a new one (Gill & Sommerville, 2023; Surian et al., 2018; cf., Pepe et al., 2025; Thomas, 2024).

Together, the work reviewed above is consistent with the possibility that when infants and toddlers see an actor behave antisocially or prosocially, they attribute some underlying stable mental state, disposition, or trait to the actor, and this attribution guides infants' and toddlers' subsequent expectations for how the actor will behave, sometimes even toward new individuals. These capacities may serve as a foundation of adults' capacities for impression formation.

Updating Impressions in Infancy

Whereas the studies reviewed above examined infants' expectations for consistency in prosocial and antisocial actions, research on early social evaluation has almost exclusively presented infants with consistent actors who act consistently, either only prosocially or only antisocially. Although infants may expect consistency, their everyday lives likely include actors that are at least sometimes inconsistent. How do infants evaluate these inconsistent individuals? To date, we are aware of only two papers have examined this question.

In one paper, Steckler et al. (2017) examined whether 9-month-olds (i) preferred reaching for an actor who consistently helped (i.e., a Consistent Helper) over an inconsistent actor who sometimes helped and sometimes hindered; and (ii) preferred reaching for an inconsistent actor who sometimes helped and sometimes hindered over an actor who consistently hindered (i.e., a Consistent Hinderer). To maximize the contrast between actors, the inconsistent actor mostly acted differently from the consistent actor: Infants either saw the consistent actor help three times and the inconsistent actor help once and hinder twice, or the consistent actor hinder three times and the inconsistent actor hinder once and help twice. Despite these and several other attempts to make the contrast between actors as easy as possible, and an independent, successful replication of infants' preference for consistent helpers over consistent hinderers, 9-month-olds consistently failed to prefer actors who helped relatively more vs. less when anyone behaved inconsistently.

In a related study, Shimizu and colleagues (2018) habituated 6-, 9-, and 12-month-old infants and 15- to 18-month-old toddlers to two actors, one who helped and the other who hindered a protagonist in its failed attempts to open a box. After the habituation phase, one of the actors then both helped *and* hindered the same protagonist to retrieve its lost ball; that is, the actor sometimes acted consistently and sometimes inconsistently with the valence of its previous

act. As reviewed above, both infants and toddlers appeared to *expect* the actor to behave consistently toward the same target (that is, infants looked longer after the inconsistently-valenced act). However, infants appeared not to use the inconsistency to inform their evaluations: Only 15- to 18-month-old toddlers subsequently preferred reaching for the more prosocial actor, whereas 6-, 9-, and 12-month-olds all chose randomly. This pattern of results is consistent with the findings with 9-month-olds from Steckler and colleagues (2017), and suggests that infants may have particular trouble generating evaluations of individuals who behave inconsistently.

The Present Experiments

Alternatively, perhaps infants actually *can* update impressions to evaluate inconsistent actors, but there were methodological aspects of the methods in past work that prevented them from demonstrating this ability. At least two aspects of the methods of past work exploring infants' ability to evaluate inconsistent actors may have prevented them from succeeding. First, the prosocial and antisocial actions used to establish infants' impressions may not have been ideal. Indeed, both Steckler et al. (2017) and Shimizu et al. (2018) primarily utilized the "box scenario," in which a protagonist tries but fails to open a box and is helped and hindered in doing so (Hamlin & Wynn, 2011). A meta-analysis on infants' social evaluations published around the same time as these papers (Margoni & Surian, 2018) suggested that infants' tendency to choose helpers over hinderers is weakest in this box-opening scenario, which might account for infants' failures to evaluate inconsistent actors in studies using it. Perhaps infants could successfully update their impressions of prosocial and antisocial actors if a different scenario were used.

Additionally, at least in Steckler et al. (2017), infants may not have been given sufficient evidence of actors' *usual* behavior to be able to form a stable impression. That is, infants were

shown the inconsistent actor acting only twice in one way; this relatively weak evidence may have rendered a weak impression (particularly if box-scenario displays are somewhat difficult to understand; Margoni & Surian, 2018). Without a stable impression of the inconsistent actor, infants may have failed to evaluate it.

The infant action understanding literature provides evidence consistent with this possibility: that just two trials' worth of evidence may be insufficient for infants to form a stable impression. Luo et al. (2017) found that when 7- to 9-month-olds saw an actor reach for one of two objects (e.g., a toy bird over a toy dog) over three familiarization trials, they subsequently looked longer during test trials when the objects switched locations and the actor reached for a different object (e.g., the dog) versus the same object (e.g., the bird). This pattern suggested that infants had attributed a goal or preference to the actor in familiarization, and expected the actor to continue acting on the same object at test (see Woodward, 1998). Crucially, when another group of 7- to 9-month-olds instead saw the actor behave inconsistently during familiarization (i.e., reaching for the bird for three trials and then the dog for one trial), they did not look longer during test when the objects switched locations and the actor reached to the relatively less-contacted object (the dog), suggestive that they had not attributed a stable goal or preference to the actor based on the actor's initial reaches. These findings suggest that, at least for infants' understanding of object-directed action, three trials of consistent reaching for one object over another is insufficient for infants to overcome subsequent (even quite minimal) inconsistency in terms of reaching for the other object. Studies of infants' expectations of consistency have presented infants with at least two to three familiarization trials in which an agent behaves one way (e.g., helping), before that agent's behavior changes in the test trials (Gill & Sommerville, 2023; Pepe et al., 2025; Surian et al., 2018; see also, Shimizu et al., 2018).

The current studies were designed to further investigate infants' ability to incorporate behavioral inconsistency into their social evaluations, while overcoming the possible explanations for infants' failures in past work raised above, including the use of a difficult scenario and too little evidence of the consistent act. The present studies adopt the "ball scenario" from Hamlin & Wynn (2011), in which a protagonist loses its ball and other actors either return the ball or take it away; this scenario was chosen because the Margoni and Surian meta-analysis (2018) suggests it has the largest effect size (see also, Tan, 2024). Additionally, Experiment 1 depicts inconsistent actors acting the same way 4 times in a row prior to any switching of behavior, in order to more clearly establish how all actors usually behave. This design allowed us to assess whether or not infants looked longer following the behavioral change, suggesting they expected the characters to behave consistently.

We conducted three experiments to probe infants' abilities to evaluate inconsistent actors. In a preliminary experiment, we first attempted to conceptually replicate the finding that infants prefer helpers over hinderers in the ball scenario from Hamlin and Wynn (2011). Following successful replication, we tested infants' evaluation of inconsistent actors in the ball scenario in Experiments 1 and 2, and we examined infants' expectations of consistency in behavior in Experiment 1.

Preliminary Experiment: Evaluations of Consistent Actors (Replication of Past Research)

We first sought to attempt to replicate the finding that, when individuals behave consistently, infants prefer actors who help others over actors who hinder others. We assessed infants' preferences using preferential looking, in which infants' preference between two entities is inferred from their relative looking to each over a set time period, following other research using this measure to probe infants' and toddlers' social evaluations (e.g., Geraci et al., 2022;

Geraci & Surian, 2023; Hamlin et al., 2010; Hamlin & Wynn, 2011; Powell & Spelke, 2018; Woo et al., 2024; Woo & Spelke, 2023b, 2023a). To our knowledge, infants' basic preference for helpers over hinderers in the ball scenario had not been previously demonstrated using online testing, and so it was important for us to examine it prior to addressing our main research question.

Methods

Hypotheses, methods, and analyses for the present experiments were preregistered on the Open Science Framework. The stimuli, data, and code are hosted at https://osf.io/879ys/?view_only=8d998b53eec148a69cbc732fc6fcb733, and the preregistration for the present experiment is hosted at https://osf.io/n7qmp/?view_only=2fa58b6c28da45c2b3a4abddac045562.

Participants

Twenty-four 11-month-olds (8 girls; mean age = 11 months, 27 days, range = 10;16 – 11;17) participated. An additional eight infants began the experiment but were not included in the final sample due to technical issues (n = 4), fussiness (n = 2), procedural error (n = 1), and inattentiveness (n = 1). Parent or guardian consent was obtained prior to participation.

Sample Size Justification. We used both pilot data (collected with the present methods) and previously published data (Hamlin & Wynn, 2011) for power analyses to determine an appropriate sample size. In pilot data (n = 8), we examined the amount of time that infants looked at the helper over the hinderer. On average, infants looked to the helper for 12.94 s and to the hinderer for 10.02 s. We ran a power analysis to determine what the sample size should be to detect a looking preference of the same effect size with sufficient power in the mixed-effects

model detailed above. With 24 infants, we would have 86% power to detect a looking preference of the same effect size.

Additionally, we ran power analyses based on published data by Hamlin and Wynn (2011) using the present method, but in a lab setting. There, 10/12 infants preferred the helper over the hinderer. We found that with 24 infants, we would have 94% power to detect a reaching preference of the same effect size. Although we are not assessing reaching preferences here, we assumed that looking and reaching preferences should be consistent with each other and should both reflect infants' evaluations (see, e.g., Hamlin et al., 2010). We therefore decided to test 24 infants.

Procedures

Infants viewed events from their parent's lap or from a high chair, from their own homes. Parents were instructed to sit quietly, look away, and not attempt to influence their infants' attention. The experimenter displayed the events on the parent's device (e.g., a laptop) over a video call, while sharing screens with the parent.

Helping and hindering events (as in Hamlin & Wynn, 2011). All events began identically. A curtain rose to reveal a green ball at the center of a stage. Two gray rabbits, wearing different shirts (pink vs. yellow), sat at the stage's corners. A protagonist (a brown bear) picked up the ball. It jumped twice, and on a third jump it dropped and retrieved the ball; it repeated this sequence three times. When the protagonist jumped after the third sequence, the ball fell toward one side and the rabbit there intervened. During helping events, the rabbit grabbed the ball. The protagonist turned to the rabbit, opening its arms as though to ask for the ball back. The rabbit turned toward the protagonist, and the two puppets faced forward. The

protagonist then turned and opened its arms again; the rabbit turned, and both faced forward again. On the protagonist's third turn, the rabbit rolled the ball toward the protagonist, who caught it. The rabbit ran off stage, and the protagonist faced forward, holding the ball. Hindering events were like helping events, except that on the protagonist's third turn, the rabbit ran off-stage with the ball. The protagonist then faced forward, empty-handed.

During all events, action paused once the acting rabbit was off-stage. Infants' looking was recorded from this point by a coder who was blind to condition using the coding program jHab (Casstevens, 2007), continuing until infants looked away for 2 consecutive seconds or 30 seconds elapsed. Infants saw a total of 8 trials, with 4 trials from each actor in alternation (Fig. 1A). One actor, the Helper, always helped the bear by returning the ball, and the other actor, the Hinderer, always took the ball away. A second coder recoded the familiarization trials of 25% of participants from this preliminary experiment and Experiments 1 and 2, and we found that the intraclass correlation coefficient for the two coders' times was 96% (95% CI [0.95, 0.97]).

Choice. Following the trials, an experimenter presented the infants with a video of the puppets. The two rabbits appeared on opposite sides of the screen and a prerecorded voice played saying "Hi! Look! Who do you like?" three times, once every 10 s, over a 30 s period, as the rabbits' hands moved as though they were clapping. A coder coded all looking at the actors during the choice period. A second coder recoded the videos of 25% of participants from this preliminary experiment and Experiments 1 and 2, and we found that the intraclass correlation coefficient for the two coders' times was 93% (95% CI [0.89, 0.95]).

Counterbalancing. We counterbalanced: (i) the color of the Helper (pink/yellow), (ii) the side of the Helper across the events and choice phase, and (iii) the order of the Helper (first/second).

Figure 1. Stimuli or the design of the replication experiment (A), and infants' preferential looking (B). For panel B, the colored boxes below the box plots indicate the behaviors that an actor engaged in (helping in orange, hindering in blue) across its four trials. White circles indicate means, horizontal lines within boxes indicate medians, pairs of connected dots indicate data from a single child, and boxes indicate interquartile ranges. The beta coefficient (β) is a measure of standardized effect size (*** $p < .001$, two-tailed).

Results

We ran a preregistered mixed-effects model on the data in the choice phase. We found that a normal distribution fit the choice data better than did a gamma distribution in the present experiment; we therefore used standard mixed-effects models for analyses of the choice phase. In this model, looking time was the dependent variable, and the actor type (Helper vs. Hinderer) was the fixed effect, with subject ID as a random effect. We found that infants looked longer in the choice phase to the Helper (mean = 10.06 s, *SD* = 3.12 s) than to the Hinderer (mean = 6.58

s, SD = 2.97 s); β = 1.00, 95% CI of β [0.61, 1.39], b = 3.47, $t(24)$ = 5.15, $p < .001$. In exploratory analyses, we categorized infants as having looked more at either the Helper or the Hinderer, and we found that 22 of 24 infants looked more at the Helper (binomial $p < .001$; Cohen's g = 0.41).

Discussion

The preliminary experiment replicated the finding that infants prefer individuals who help others over individuals who hinder others, using a different dependent variable than past research with the same stimuli. Having replicated this finding via online testing and preferential looking, we next moved onto the key aim of the research: probing infants' evaluations of inconsistent actors.

Experiment 1: Evaluations of Inconsistent Actors

In Experiment 1, an Inconsistent Actor either (i) helped a protagonist (giving its ball back) 4 times before hindering the protagonist (taking its ball away) in 1 last trial (Inconsistent Helper Condition) or (ii) hindered the protagonist 4 times before helping the protagonist in 1 last trial (Inconsistent Hinderer Condition). As in Steckler et al. (2017), the Consistent Actor always acted consistently, either helping or hindering; the Inconsistent Actor was thus maximally different from the Consistent Actor. Infants then chose between the Consistent and Inconsistent Actors.

If 11-month-olds incorporate inconsistency into their social evaluations using something like an averaging process, then they should prefer the actor who helps relatively more vs. less, irrespective of consistency and valence (see Fig. 2). Specifically, when the Consistent Actor always helps and the Inconsistent Actor mostly hinders, infants should prefer the Consistent Actor. When the Consistent Actor always hinders and the Inconsistent Actor mostly helps,

infants should prefer the Inconsistent Actor. Alternatively, infants' preferences may reflect some bias, following the adult literature. For example, infants' preferences could reflect a negativity bias, whereby infants discount positive information (i.e., preferring a Consistent Actor who only helps over an Inconsistent Actor who mostly hinders, but not distinguishing between a Consistent Actor who only hinders and an Inconsistent Actor who mostly helps) (Baumeister et al., 2001; Ferguson et al., 2019; Kanouse & Hanson Jr., 1987; Rozin & Royzman, 2001; Skowronski & Carlston, 1989). Finally, infants may not have a clear preference for the Consistent vs. Inconsistent Actor in any condition, suggesting that the adjustments made relative to Steckler et al. (2017) were not sufficient to allow infants to incorporate inconsistency into their evaluations.

In addition to examining infants' preferential looking, we also examined whether they had expectations for consistency. If so, then they should look longer when the Inconsistent Actor's behavior changes, relative to their looking in the equivalent trials for the Consistent Actor.

Methods

The preregistrations for Experiments 1 and 2 are hosted at https://osf.io/bjy5x/?view_only=c0094a1ac03441cc83811872ddc8d35c.

Participants

Forty-eight 11-month-olds (21 girls; mean age = 10 months, 26 days; range = 10;14 – 11;26) participated. Infants were randomly assigned to the "Inconsistent Hinderer" Condition, wherein the inconsistent actor hindered 4 times and helped once, or the "Inconsistent Helper" Condition, wherein the inconsistent actor helped 4 times and hindered once (n = 24 per

condition). An additional four infants began the experiment but were not included in the final sample due to technical issues (n = 2), fussiness (n = 1), and inattentiveness (n = 1).

Figure 2. A conceptual schematic for what infants have observed in past research (and in our preliminary replication experiment) and what infants observed in the present research (A), the representations that infants would have of those actors (B), and the preferences that infants may

have based on those representations (B). Together, our experiments test for several key strategies that infants may use when processing inconsistency (B). Across panels, orange squares indicate acts of helping, blue squares indicate acts of hindering, white curved rectangles indicate an actor who behaved consistently, and gray curved rectangles indicate an actor who behaved inconsistently.

Sample Size Justification. We used pilot data (collected with the methods for Experiment 1) for power analyses in order to determine an appropriate sample size. In pilot data (n = 9), we examined the amount of time that infants looked at the more prosocial over the more antisocial actor by condition. When the Inconsistent Actor was usually antisocial, infants looked at the Inconsistent Actor (here, the more antisocial actor) for 8.53 s and the Consistent Actor for 9.75 s. When the Inconsistent Actor was usually prosocial, infants looked at the Consistent Actor (here, the more antisocial actor) for 6.65 s and the Inconsistent Actor for 13.14 s. The effect appeared qualitatively larger, then, when the Inconsistent Actor was usually prosocial; however, there was a very small amount of data per condition.

Nonetheless, we ran a power analysis to determine what the sample size should be to detect a looking preference between the Consistent and Inconsistent Actors of the same effect size with sufficient power as in the mixed-effects model detailed above. This analysis suggested that there was an interaction, such that the effect was larger when the Inconsistent Actor was usually prosocial, consistent with the qualitative trend above. With 48 infants (24 per condition), we would have 100% power to detect this interaction. We also ran the same analysis, but without the Inconsistent Actor's usual act in the model. With 48 infants (24 per condition), we would

have 99% power to detect a looking preference for the more prosocial actor. Thus, our overall analysis is sufficiently powered whether the interaction between conditions observed during piloting was stable.

In the pilot data, we found almost no evidence in our online pilot that infants looked longer when the Inconsistent Actor's behavior changed. Overall, in the trial before the Inconsistent Actor's behavior changed, infants looked 13.30 s. In the trial involving the change, infants looked 9.64 s. Given that infants appeared to not look longer following this change, and given the uncertainty of the online testing, we chose to focus our power analyses on infants' preferential looking in the choice test. However, we still planned to examine infants' expectations of the actors' behaviors in the larger sample.

Figure 3. Representative stimuli for the designs of Experiment 1. Orange indicates helping or prosocial behavior, blue indicates hindering or antisocial behavior, green indicates the fourth trial for an actor (consistent for both Actors), and purple indicates the fifth trial for an actor (inconsistent only for the Inconsistent Actor). The screenshots are representative of Experiment

1's the "Inconsistent Actor Mostly Hinders" Condition; the "Inconsistent Actor Mostly Helps" Condition is not shown.

Procedures

The procedure was like that of the preliminary replication experiment, except that there was an additional, fifth trial for each actor. Thus, infants saw a total of 10 trials. We chose 5 trials per actor so that there would be the same evidence of consistent behavior (4 trials) prior to any inconsistency, given that the replication experiment suggested that 4 trials was sufficient for infants to generate evaluations in the current context (see Fig. 3).

Consistent Actors performed the same action in all 5 of their trials (either always returning the ball, or always taking the ball away), whereas Inconsistent Actors would act opposite to the Consistent Actor for their first 4 trials, and then would switch its behavior to behave similarly to the Consistent Actor on the last trial. To reduce working memory demands, each actor completed its trials consecutively (instead of on alternating trials; see discussion in Steckler et al., 2017). That is, one actor acted in all 5 of its trials before the other began acting. Infants were randomly assigned to the "Inconsistent Hinderer" Condition, in which the Inconsistent Actor mostly hindered and the Consistent Actor helped; or the "Inconsistent Helper" Condition, in which the Inconsistent Actor mostly helped and the Consistent Actor hindered.

Coding was identical to that of the preliminary replication experiment.

Counterbalancing. We counterbalanced: (1) the color of the Consistent Actor (pink/yellow); (2) the side of the Consistent Actor across the events and choice phase (left/right); and (3) the order of the Consistent Actor (first/second).

Figure 4. Infants' preferential looking (A) and expectations (B) in Experiment 1. Blue indicates hindering or antisocial behavior, orange indicates helping or prosocial behavior, green indicates the fourth trial for an actor (consistent for both actors), and purple indicates the fifth trial for an actor (inconsistent only for the Inconsistent Actor). In panel A, the colored boxes below box plots indicate the behaviors that an actor engaged in (helping in blue, hindering in orange) across its five trials; and the colored boxes in gray indicate that that actor was inconsistent. In the box plots of panel B, white circles indicate means, horizontal lines within boxes indicate medians, pairs of connected dots indicate data from a single child, and boxes indicate interquartile ranges. The data in panel B are collapsed across actor conditions. Beta coefficients (β) are measures of standardized effect size, whereas regression coefficients (*b*) for the gamma models are non-standardized (NS $p > .05$; * $p < .05$, two-tailed). (Standardization is not possible for the gamma model, because it requires all values are 0 or larger.)

Results

Choice Phase

We first examined whether infants preferred the actor who was on average more prosocial. We ran a preregistered mixed-effects model on the data in the choice phase. In this model, looking time was the dependent variable, with fixed effects of actor type (More Prosocial vs. the More Antisocial Actor; within-subjects), condition (Inconsistent Helper/Inconsistent Hinderer; between-subjects), and their interaction. Subject ID was a random effect. We centered the fixed effects. Analyses revealed only a main effect of actor type: Infants looked longer to the More Prosocial Actor (mean = 12.01 s, *SD* = 5.63 s) than to the More Antisocial Actor (mean = 9.64 s, *SD* = 4.98 s) (β = 0.44, 95% CI of β [0.05, 0.83], *b* = 2.37, *t*(96) = 2.22, *p* = .028; see Fig. 4A). The non-significant interaction between actor type and condition (*p* = .228) indicated that infants were no more likely to prefer the More Prosocial Actor versus the More Antisocial Actor when either the Helper or Hinderer acted consistently. In exploratory analyses, we categorized infants as having looked more at either the More Prosocial Actor or the More Antisocial Actor, and we found that 27 of 48 infants looked more at the More Prosocial Actor (binomial *p* = .470; Cohen's *g* = 0.06).

Expectations of Consistency

Next, we asked whether infants formed expectations for consistency, as well as whether expectations varied by valence. Because looking time data are often positively skewed, we first fit both a gamma distribution and a normal distribution to the data. We found that a gamma distribution fit the data better than did a normal distribution; we therefore ran a gamma mixed-effects model. Looking time was the dependent variable, with actor type (Consistent/Inconsistent Actor; within-subjects), trial order (corresponding with the trials involving the switch in behavior (for Inconsistent Actors; within-subjects) and the one before that), the Inconsistent Actor's usual act (Helping/Hindering; between-subjects), and all possible interactions as fixed effects.

Participant ID was a random effect. The fixed effects of actor type, order, and usual act were centered.

Analyses revealed no significant interactions (two-way or three-way) or main effects (all *p*s > .18). For both conditions, the change in infants' looking between the fourth and fifth trials was not significantly different for the Consistent Actor ($mean_{Trial\ 4}$ = 10.21 s; $SD_{Trial\ 4}$ = 7.36 s; $mean_{Trial\ 5}$ = 10.40 s, $SD_{Trial\ 5}$ = 7.03 s) vs. the Inconsistent Actor ($mean_{Trial\ 4}$ = 9.57 s; $SD_{Trial\ 4}$ = 7.63 s; $mean_{Trial\ 5}$ = 11.10 s, $SD_{Trial\ 5}$ = 7.58 s) (β = 0.23, 95% CI of β [-0.11, 0.57], *b* = 0.22, *t* = 1.32, *p* = .184; see Fig. 4B). That is, infants did not appear to notice when the Inconsistent Actor changed its behavior. Indeed, although infants looked approximately 1 second longer to the Inconsistent Actor's inconsistent versus its consistent event, this difference did not reach significance, nor was it significantly different from the same difference for the Consistent Actor.

Discussion

In Experiment 1, we found that infants appeared to successfully evaluate inconsistent actors: They preferentially looked at the More Prosocial Actor in each condition, despite the fact that one of the actors behaved inconsistently. We note that there was evidence that infants formed preferences in Experiment 1, this evidence was qualitatively weaker than in the preliminary experiment in which actors only behaved consistently, with there being a smaller looking time difference (with β = 1.00 in the preliminary experiment and β = 0.44 in Experiment 1) and the number of infants looking more to the More Prosocial Actor not differing from chance in Experiment 1.

Although the evidence was weak, infants' preferences in Experiment 1 stand in contrast with past research that has presented infants with inconsistent actors, in which both Steckler and

colleagues (2017) and Shimizu and colleagues (2018) found that infants from 6-12 months of age failed to preferentially reach for actors who were on average more prosocial.

Why might infants have evaluated inconsistent actors in Experiment 1, but not in past research? One possibility is that inconsistency is indeed difficult for infants to evaluate, and past work included additional inconsistencies that made the difficulty too great; these additional inconsistencies were eliminated in the current studies. For instance, Steckler et al. (2017) used live puppet shows, which, given they are performed by human experimenters, inevitably vary somewhat from one trial to the next even when the nature of the unfulfilled goal scenario and the valence of the action remains consistent. In contrast, the current Experiment 1 utilized prerecorded videos of puppet shows that were entirely identical across trials, minimizing any extraneous inconsistencies. Likewise, Shimizu et al. (2018)'s inconsistent acts were from a novel scenario involving an entirely new unfulfilled goal relative to what infants had seen in habituation (retrieving a lost ball rather than opening a box); this inconsistency in scenario may have proven too significant for infants to maintain their evaluations of actors who were also inconsistently prosocial/antisocial over time. In contrast, Experiment 1 presented infants with the same scenario throughout (always involving retrieving a lost ball); only the valence of the actor's behavior (sometimes) changed. Together, the combination of Experiment 1 with past studies suggest that infants can only incorporate inconsistency into their evaluations if other forms of inconsistency are minimized.

That said, another possibility is that infants only appeared to successfully incorporate inconsistency into their evaluations in Experiment 1 but did not actually do so. Indeed, although infants' attention did increase somewhat in response to the Inconsistent Actor's inconsistent behavior as in past work (e.g., Shimizu et al., 2018; see also, Taborda-Osorio et al., 2019; Tatone

et al., 2015), the effect was not significant nor significantly different from their looking to the Consistent Actor's behavior on the same trial. These null findings suggest that the present infants may not have noticed the change in the Inconsistent Actor's behavior. We note that the lack of a significant finding may reflect that our study was underpowered to detect differences: Given our focus on social evaluation, our power analyses from pilot data were based on attention during the preferential looking trial rather than on event-based attention differences. That said, the lack of significant effect of attention on behavioral inconsistency means that this group of infants may have succeeded on the preferential looking trial simply because they failed to notice the inconsistent acts in the first place. If so, it would be inappropriate to argue that infants incorporated inconsistency into their evaluations.

Regardless of whether infants noticed the inconsistency, it is possible that initial consistency supports infants' evaluations of inconsistent actors. In Experiment 1, the Inconsistent Actor always behaved in a way that was opposite that of the Consistent Actor, with the Inconsistent Actor switching its behavior in the last trial. Infants in the present experiment may have evaluated the Inconsistent Actor on the basis of its initial consistency. Evidence for this possibility comes from research on goal attribution. In classic research on goal attribution, Woodward (1998) found that after 6- and 9-month-old infants observe a hand grasping for one toy over another repeatedly, infants look longer when the hand later grasps the other, previously ignored toy, than when the hand later grasps the original object, as though infants had expected the hand to continue reaching for the same object. Building on Woodward's findings, Duh and Wang (2019) examined 14-month-olds' abilities to form such expectations from inconsistent behavior. They found that 14-month-olds can attribute goals to a person who behaves inconsistently, but only if the person's initial behavior is consistent: 3 consecutive actions on the

same object, preceding a switch in the target of the person's actions. But when the person's initial behavior was inconsistent (i.e., acting on two different objects in the first three trials), even when infants observed the same number of actions on an object, they did not attribute a goal to the person. It is possible that infants in the present work must also initially observe consistent behavior to evaluate inconsistent actors. We therefore designed Experiment 2 to attempt to conceptually replicate the findings of Experiment 1 while addressing these concerns about the timing of inconsistency.

Experiment 2: Evaluations of Inconsistent Actors (Online)

To conceptually replicate the preferential looking findings of Experiment 1 and address the concerns about the timing of the inconsistency in Experiment 1, in Experiment 2 the Inconsistent Actor performed its unusual act first, before switching its behavior and acting in the opposite way 4 times in a row. If the preferential looking findings from Experiment 1 resulted from infants ignoring the final inconsistent act and/or focusing entirely on what happened first, or if infants require evidence of consistency over at least three trials before they can make sense of inconsistency (as in Duh & Wang, 2019), then infants should not prefer the actor who is on average more helpful in Experiment 2. But if infants indeed incorporate inconsistency into their evaluations by averaging across all observed behaviors, regardless of the timing of inconsistency, then infants should still prefer the actor who is on average more helpful in Experiment 2.

Methods

Participants

Sixty-nine 11-month-olds (42 girls; mean age = 10 months, 29 days; range = 10;15 – 11;20) participated. Infants were randomly assigned to the "Inconsistent Actor Mostly Hinders" Condition (n = 35) or the "Inconsistent Actor Mostly Helps" Condition (n = 34).

Sample Size Justification. We used previously collected data (from Experiment 1 and a pilot study) for power analyses in order to determine an appropriate sample size. In these data (n = 57), we examined the amount of time that infants looked at the more prosocial over the less prosocial actor by condition. Infants preferred looking at the more prosocial actor.

We ran a power analysis to determine what the sample size should be to detect a looking preference of the same effect size with sufficient power in the mixed-effects model detailed above. With 64 infants (32 per condition), we would have 83% power to detect a looking preference for the more prosocial actor. More families responded to recruitment efforts than expected, leaving us with a total of 69 infants.

Procedures

The procedure was like that of Experiment 1, except that the Inconsistent Actor's unusual act took place in its first trial (see Fig. 4A).

Results

Choice Phase

Preregistered Analyses. As in Experiment 1, we examined whether infants preferred the actor who was on average more prosocial. We ran a preregistered mixed-effects model on the data in the choice phase using the same model as in Experiment 1. None of the main effects and interactions were significant (all ps > .41; Fig. 4B). Infants did not differ in their looking to the More Prosocial Actor (mean = 8.79 s, SD = 5.20 s) and to the More Antisocial Actor (mean = 7.99 s, SD = 5.20 s) during the preferential looking trial.

Exploratory Analyses. In exploratory analyses, we categorized infants as having looked more at either the More Prosocial Actor or the More Antisocial Actor, and we found that 35 of 69 infants looked more at the More Prosocial Actor (binomial $p = 1.00$; Cohen's $g = 0.00$).

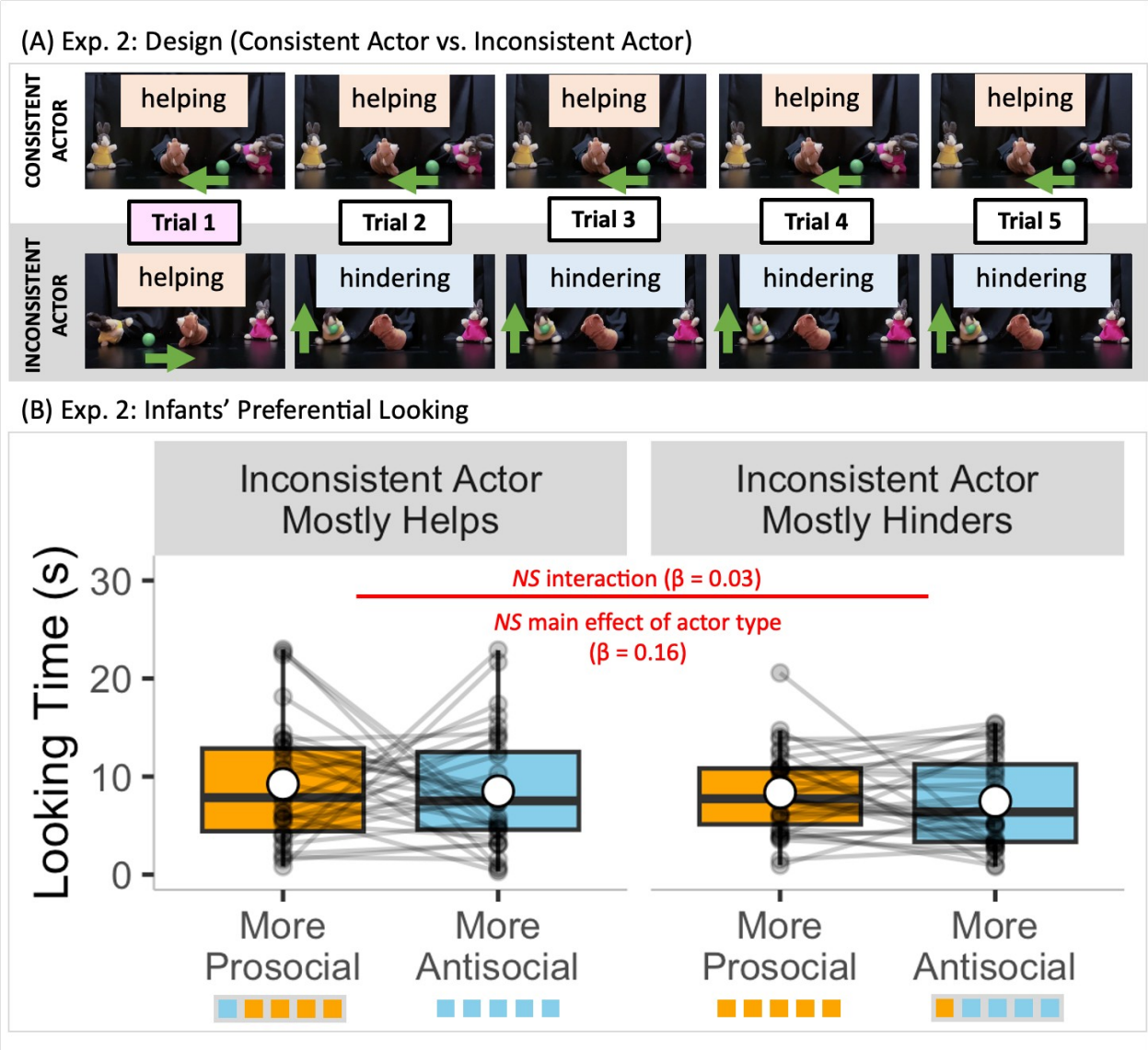

Figure 4. Stimuli for Experiment 2 (A) and infants' preferential looking (B). Across panels, blue indicates hindering or antisocial behavior, orange indicates helping or prosocial behavior, and

purple indicates the trial for which the Inconsistent Actor's behavior was different than it was for its other trials. In panel B, the colored boxes below the box plots indicate the behaviors that an actor engaged in (helping in blue, hindering in orange) across its five trials; and the colored boxes in gray indicate that that actor was inconsistent. White circles indicate means, horizontal lines within boxes indicate medians, pairs of connected dots indicate data from a single child, and boxes indicate interquartile ranges. Beta coefficients (β) are measures of standardized effect size (NS $p > .05$).

Given that the change in behavior occurred between trials 1 and 2, before collecting data, we did not plan to examine expectations of consistency. In exploratory analyses, we examined whether infants' attention changed between trials 1 and 2, depending on whether the actor's behavior had or had not changed. This model was like that of Experiment 1, except that the trials were 1 and 2, rather than 4 and 5.

We first fit both a gamma distribution and a normal distribution to the data. We found that a gamma distribution fit the data better than did a normal distribution. We therefore ran a gamma mixed-effects model. There was a significant main effect of trial order, such that infants looked less in the second trial (mean = 11.43 s, SD = 7.93) than in the first trial (mean = 13.59 s, SD = 8.02 s) for a given actor. However, the three-way interaction between actor type, trial order, and the usual act of the Inconsistent Actor was not significant ($p = 0.391$), nor was the two-way interaction between trial order and actor type ($p = 0.152$). That is, the effect of trial order did not differ depending on whether the actor was Inconsistent or Consistent, in either of the two conditions. We do not interpret these findings as evidence of infants' expectations, given

that infants only had one trial of evidence before the switch in behavior, and infants may not have formed strong expectations from that one trial.

Discussion

In Experiment 2, for which the Inconsistent Actor's unusual act occurred first, infants did not incorporate inconsistency into their evaluations. In both conditions, they did not prefer the actor who had been on average more prosocial. These null findings suggest that infants do not reliably incorporate inconsistency into their evaluations. These findings stand in contrast with the positive findings that infants evaluate inconsistent actors in Experiment 1, for which the Inconsistent Actor's unusual act occurred on its last trial. Together, these findings suggest that infants may not incorporate inconsistency in their evaluations; they may fail to notice it and instead evaluate actors based on their initial behaviors.

General Discussion

In the present experiments, we replicated the finding that infants prefer individuals who are helpful over agents who are unhelpful using an online testing platform, and we found some evidence that 11-month-old infants can evaluate individuals who behave inconsistently using preferential looking as a dependent variable (Experiment 1). However, we did not conceptually replicate this evidence in Experiment 2, and we found no evidence that infants noticed inconsistency in behavior. Together, these findings suggest that although infants readily distinguish between consistently prosocial and antisocial actors, that they do not reliably incorporate inconsistency into their evaluations.

Capacities to Evaluate Inconsistent Actors in the First Year

Experiment 1's findings that 11-month-old infants can evaluate inconsistent actors stands in contrast with past research that has found that infants in the first year struggle to evaluate

actors who behave inconsistently (Shimizu et al., 2018; Steckler et al., 2017). However, in the present experiments, infants only succeeded at evaluating inconsistent actors when the Inconsistent Actor and the Consistent Actor initially behaved differently (as in Experiment 1); infants' success was weak in Experiment 1; and infants failed when the two actors initially behaved the same (as in Experiment 2). We take these findings as evidence that, at 11 months of age, infants do not engage in holistic impression formation: that they do not incorporate inconsistency into their social evaluations through an averaging process.

Instead, infants may ignore the inconsistency: They may follow a primacy bias, prioritizing their first impressions of the actors (i.e., how the actors initially behaved). This possibility may also explain the findings of Shimizu et al. (2018); although Shimizu et al. found that 15- to 18-month-old toddlers (but not younger infants) evaluated inconsistent actors, Shimizu et al. only ever showed participants stimuli in which the Inconsistent Actor initially behaved differently from the Consistent Actor. Thus, toddlers may have evaluated the actors based on their initially different behaviors. The present findings only provide tentative evidence of a primacy bias, and further studies would be necessary to examine whether a primacy bias guides infants' social evaluations.

In the present experiments, it is further possible that infants may not have even noticed the inconsistency. In Experiment 1, infants did not look significantly longer when the Inconsistent Actor's behavior changed; however, the data trended in that direction, and it is unclear if the experiment was underpowered to detect that effect. Regardless of whether infants noticed the inconsistency in the present experiments, it appears that inconsistency is difficult for infants to incorporate in their social evaluation.

Why is inconsistency so difficult for infants? One possibility is that infants may just not be very good at social evaluation, and so any inconsistency disrupts their ability to evaluate prosocial versus antisocial actors at all. Indeed, although we successfully replicated infants' preference for consistent prosocial versus antisocial actors in the current studies, there are numerous failed replications of infant social evaluation effects in the literature (Lucca, Yuen, et al., 2025; Salvadori et al., 2015); these failures may indicate that early capacities for social evaluation are simply fragile. If infants' evaluative capacities are indeed fragile, then behavioral inconsistency may prove challenging in that it overwhelms an already weak system.

That said, numerous other studies have shown that infants' and toddlers' evaluations of consistent actors can be complex, incorporating variation in actors' intentions, knowledge, and even beliefs (Geraci et al., 2022; Geraci & Surian, 2023; Hamlin, 2013a; Kanakogi et al., 2017; Woo et al., 2017; Woo & Spelke, 2023b). Interestingly, infants and toddlers appear to not be overtaxed in these studies, wherein the constructs being tested are complex, but no actor behaves inconsistently.

It is possible that infants may hold a basic assumption that people's actions are based on stable underlying mental states, and should therefore remain consistent over time (for related findings with adults, see Anderson, 1965; Anderson & Hubert, 1963; Kim et al., 2020). Evidence in the goal attribution literature (Duh & Wang, 2019) suggests that infants need to initially observe an actor behaving consistently (as in Experiment 1), before they can process inconsistent behavior. That is, infants may be able to engage in holistic impression formation, but their abilities may be limited, requiring that they first observe an actor engaging in consistent behavior. If infants, like adults, generally assume that people will behave consistently, the violation of this assumption may disrupt their abilities to evaluate agents, leading them to ignore

inconsistent information and, when possible, rely on first impressions based on an actor's initially consistent behavior.

Conclusion

Although we found evidence in one experiment that 11-month-old infants evaluated inconsistent actors, we did not conceptually replicate that evidence in a further experiment, and they did not appear to notice inconsistency in behavior in any experiment. The present experiments contribute to a growing body of research that infants can experience difficulty evaluating inconsistent actors, even when the inconsistency is very low (1 of 5 acts) and even when a separate group of infants succeeded in evaluating characters who consistently performed the very same acts.

Evaluating inconsistent behavior is a challenge that older children and even adults have difficulty with and limited success in overcoming. Yet, the ability to incorporate behavioral inconsistency into social evaluation is critical to functioning in the real world, where people regularly behave inconsistently. The present study suggests that the limitations of this ability may have roots in infancy. We call on future research to continue testing the link between infants' and toddlers' expectations of consistency and their evaluations of inconsistent actors, to better understand how children come to incorporate inconsistency into their evaluations.

References

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*(3), 518–529. https://doi.org/10.1037/0022-3514.69.3.518

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256–274. https://doi.org/10.1037/0033-2909.111.2.256

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, *2*(1), 1.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, *2*(5–6), 379–391.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Brosch, T., Schiller, D., Mojdehbakhsh, R., Uleman, J. S., & Phelps, E. A. (2013). Neural mechanisms underlying the integration of situational information into attribution outcomes. *Social Cognitive and Affective Neuroscience*, *8*(6), 640–646.

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.

Casstevens, R. M. (2007). jHab: Java habituation software (version 1.0. 2)[computer software]. *Chevy Chase, MD*.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, *163*, 163–228.

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

Darwin, C. (2008). *The descent of man, and selection in relation to sex*. Princeton University Press.

Duh, S., & Wang, S. (2019). Infants Detect Patterns of Choices Despite Counter Evidence, but Timing of Inconsistency Matters. *Journal of Cognition and Development*, *20*(1), 96–106. https://doi.org/10.1080/15248372.2018.1528976

Ferguson, M. J., Mann, T. C., Cone, J., & Shen, X. (2019). When and How Implicit First Impressions Can Be Updated. *Current Directions in Psychological Science*, *28*(4), 331–336. https://doi.org/10.1177/0963721419835206

Geraci, A. (2022). Some considerations for the developmental origin of the principle of fairness. *Infant and Child Development*, *31*(6), e2350. https://doi.org/10.1002/icd.2350

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.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21.

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.

Hamlin, J. K. (2013a). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), Article 3. https://doi.org/10.1016/j.cognition.2013.04.004

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*(3), Article 3. https://doi.org/10.1177/0963721412470687

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., & 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

Henrich, N., & Henrich, J. P. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford University Press.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24.

Joyce, R. (2007). *The evolution of morality*. MIT press.

Kanakogi, Y., Inoue, Y., Matsuda, G., Butler, D., Hiraki, K., & Myowa-Yamakoshi, M. (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behaviour*, *1*(2), Article 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.

Kanouse, D. E., & Hanson Jr., L. R. (1987). Negativity in evaluations. In *Attribution: Perceiving the causes of behavior* (pp. 47–62). Lawrence Erlbaum Associates, Inc.

Katz, L. D. (2000). *Evolutionary origins of morality: Cross-disciplinary perspectives* (Vol. 1). Imprint Academic.

Kim, M. J., Park, B., & Young, L. (2020). The psychology of motivated versus rational impression updating. *Trends in Cognitive Sciences*, *24*(2), 101–111.

Kim, M. J., Theriault, J., Hirschfeld-Kroen, J., & Young, L. (2022). Reframing of moral dilemmas reveals an unexpected “positivity bias” in updating and attributions. *Journal of Experimental Social Psychology*, *101*, 104310.

Kubota, J. T., Mojdehbakhsh, R., Raio, C., Brosch, T., Uleman, J. S., & Phelps, E. A. (2014). Stressing the person: Legal and everyday person attributions under stress. *Biological Psychology*, *103*, 117–124.

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

Nighbor, T., Kohn, C., Normand, M., & Schlinger, H. (2017). Stability of infants' preference for prosocial others: Implications for research based on single-choice paradigms. *PloS One*, *12*(6), e0178818.

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 Conference of the Cognitive Science Society*.

Powell, L. J., & Spelke, E. S. (2018). Third-party preferences for imitators in preverbal infants. *Open Mind*, *2*(2), Article 2.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Elsevier. https://www.sciencedirect.com/science/article/pii/S0065260108603573

Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

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.

Shimizu, Y., Senzaki, S., & Uleman, J. S. (2018). The Influence of Maternal Socialization on Infants' Social Evaluation in Two Cultures. *Infancy*, *23*(5), 748–766. https://doi.org/10.1111/infa.12240

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131.

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.

Taborda-Osorio, H., Lyons, A. B., & Cheries, E. W. (2019). Examining infants' individuation of others by sociomoral disposition. *Frontiers in Psychology*, *10*, 1271.

Tan, A. W. (2024). *Prosocial evaluation: A new meta-analysis*. https://files.osf.io/v1/resources/qnh92/providers/osfstorage/663af06697d14237bbdadf0d?format=pdf&action=download&direct&version=1

Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62. https://doi.org/10.1016/j.cognition.2014.12.007

Thomas, A. J. (2024). Cognitive Representations of Social Relationships and their Developmental Origins. *Behavioral and Brain Sciences*, 1–53. https://doi.org/10.1017/S0140525X24001328

Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. *The New Unconscious*, 362–392.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237.

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. (2023a). Infants and toddlers leverage their understanding of action goals to evaluate agents who help others. *Child Development*.

Woo, B. M., & Spelke, E. (2023b). Toddlers' social evaluations of agents who act on false beliefs. In *Developmental Science*. https://doi.org/10.31234/osf.io/eczgp

Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, *168*, 154–163. https://doi.org/10.1016/j.cognition.2017.06.029

Woo, B. M., Tan, E., & Hamlin, K. (2022). Human morality is based on an early-emerging moral core. In *Annual Review of Developmental Psychology*. https://doi.org/10.31234/osf.io/98d36

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), Article 1.

Acknowledgments

We thank the families who participated in these studies; Gia Kakkar, Angela Zhu, Julian Kwong, Anni Persson, Brittany Tsang, Mehek Budshah, Fibha Khan, Caroline Mawhinney, and Chloe Fichter for research assistance; Francis Yuen for helpful discussion; and ChildrenHelpingScience.com for support with recruitment.