Cognition 167 (2017) 255–265

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The limits of early social evaluation: 9-month-olds fail to generate social evaluations of individuals who behave inconsistently

Conor M. Steckler *, Brandon M. Woo, J. Kiley Hamlin

University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 23 January 2016
Revised 11 March 2017
Accepted 25 March 2017
Available online 7 April 2017

Keywords:
Social evaluation
Infant
Moral learning
Behavioral inconsistency

ABSTRACT

Infant studies examining the development of the ability to evaluate others for their pro- and antisocial acts to date have explored how infants evaluate individuals who are either consistently prosocial or consistently antisocial. Yet in the real world, one regularly encounters individuals who behave inconsistently, engaging in multiple different kinds of behaviors that are variably prosocial and antisocial. In order to form accurate social evaluations of these inconsistently helpful and harmful individuals, then, evaluators must be able to aggregate across different types of behaviors and update previously formed evaluations based on new information. The current studies were designed to examine 9-month-old infants' social evaluations of characters who have displayed both prosocial and antisocial acts. Across three experiments using a previously utilized scenario for testing infants' preference for prosocial over antisocial others, infants repeatedly failed to prefer more- versus less-prosocial individuals when one of those individuals had previously acted both prosocially and antisocially, despite various attempts to facilitate responding across experiments. Notably, an additional experiment replicated infants' preference for consistently prosocial over consistently antisocial others. Together, findings from the current studies suggest that incorporating behavioral inconsistency into one's social evaluations may be especially difficult for infants in the first year.

1. Introduction

A basic requirement for human morality is the ability to evaluate others on the basis of how they treat third parties. Indeed, although both the scope of the moral domain and the mechanisms by which moral judgments are reached are topics of considerable empirical and theoretical debate (Blair, 1995; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Haidt, 2001; Haidt & Joseph, 2007; Kohlberg, 1969; Turiel, 1983), it seems uncontroversial that a significant proportion of human moral judgments require at least (1) identifying actions performed with the intention of helping or harming others, and (2) generating positive evaluations of those who intend to help and negative evaluations of those who intend to harm.

Research into the development of these abilities has demonstrated that humans begin showing sensitivity to certain aspects of the moral world extremely early in development, perhaps within the first months after birth (see review in Hamlin, 2013a). Specifically, by just 3 months of age infants look longer at a character who previously helped a 3rd party achieve its unfulfilled goal than at a character who previously hindered a 3rd party, and as soon as infants can reliably reach (by about 4.5 months) they selectively touch prosocial characters over antisocial ones (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007, 2010). These evaluations occur throughout infancy in response to a variety of sociomoral interactions and scenarios (Buon et al., 2014; Burns & Sommerville, 2014; Geraci & Surian, 2011; Hamlin & Wynn, 2012; Hamlin, Wynn, Bloom, & Mahajan, 2011; Scola, Holvoet, Arciszewski, & Picard, 2015; but see Salvadori et al., 2015), suggestive that infants' evaluations are relatively robust. Furthermore, infants' evaluations are based on others' prosocial and antisocial mental states as opposed to the outcomes they bring about (Hamlin, 2013b; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; see also Choi & Luo, 2015; Lee, Yun, Kim, & Song, 2015), and differ depending on the context in which prosocial and antisocial acts occur (e.g., Hamlin, 2014; Hamlin et al., 2011). These results suggest that, like adults', infants' social evaluations are sensitive to why others act as they do.

Work suggestive that human infants are precocious evaluators of the sociomoral world is consistent with theories that the human moral sense may have evolved to sustain group living (see review in Joyce, 2006). However, whereas infant studies to date have

* Corresponding author at: University of British Columbia, Department of Psychology, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada.
E-mail address: conor.steckler@psych.ubc.ca (C.M. Steckler).

http://dx.doi.org/10.1016/j.cognition.2017.03.018
0010-0277/

explored how infants evaluate individuals who engage in just one kind of behavior (behaving either prosocially *or* antisocially), in the real world it is presumably far more common to encounter individuals who engage in multiple different kinds of behaviors, ones that are variably prosocial and antisocial. Indeed, it may be more appropriate to define "good" individuals as those who behave prosocially much or most of the time, and "bad" individuals as those who more often behave antisocially. In order to form accurate social evaluations of good and bad individuals in one's everyday life, then, evaluators must be able to aggregate across different types of behaviors, and, if necessary, update previously formed evaluations based on new information. The current studies were designed to examine whether infants are able to form social evaluations of characters who are inconsistently prosocial and antisocial.

2. Forming and updating social impressions in adults and children

The process of forming and updating representations of others based on their trait-relevant behaviors has been variably referred to as person perception, impression formation, trait understanding, and/or trait attribution (see, e.g., Heider, 1958; Jones & Davis, 1965; Kelley, 1967). Adults readily evaluate and form impressions of others based on exceedingly small amounts of trait-relevant information (Ambady & Rosenthal, 1992; Carlston & Skowronski, 1994; Winter & Uleman, 1984), sometimes failing to recognize that certain behaviors may reflect situation-specific causes rather than stable dispositional traits (Gilbert & Malone, 1995; Jones & Harris, 1967; Ross, 1977). Perhaps because it is inherently uncomfortable for adults to discover information that is inconsistent with their previously established social impressions (Heider, 1958), research has shown that they use a variety of strategies to reconcile the inconsistencies they encounter (see Anderson, 1965; Abelson et al., 1968; Heider, 1958; Hendrick, 1968; Srull & Wyer, 1989). These strategies include simply ignoring inconsistent information, aggregating all of the information observed via a simple averaging function, and attempting to explain inconsistent behaviors in light of a previously established impression, amongst others.

The development of person perception in childhood has also received considerable empirical attention (e.g., Alvarez, Ruble, & Bolger, 2001; Boseovski, Chiu, & Marcovitch, 2013; Boseovski & Lee, 2006; Bretherton & Beeghly, 1982; Cain, Heyman, & Walker, 1997; Dozier, 1991; Droege & Stipek, 1993; Heller & Berndt, 1981; Hendrick, Franz, & Hoving, 1975; Liu, Gelman, & Wellman, 2007; Livesley & Bromley, 1973; Mrug & Hoza, 2007; Rholes & Ruble, 1984; Rosenbach, Crockett, & Wapner, 1973; Ruble & Dweck, 1995). Generally, this research has examined whether and when children begin to form impressions and assign trait labels to others on the basis of trait-relevant behaviors, as well as whether children are able to use others' past behaviors to predict which behaviors they are most likely to perform in the future. Although many studies have demonstrated that it is not until middle childhood that children have a strong grasp of the relationship between past behaviors, traits, and future behaviors (see Ruble & Dweck, 1995; Yuill, 1992; for reviews); other research shows various aspects of these abilities emerge during the preschool and early school years (Alvarez et al., 2001; Boseovski & Lee, 2006; Boseovski et al., 2013; Cain et al., 1997; Liu et al., 2007). Notably, young children appear to reason about some traits (e.g., social traits) more easily than others (e.g., non-social, such as academic traits); for example, they are particularly adept at generating trait attributions and behavioral predictions when an agent's behaviors were prosocial or antisocial (Benenson & Dweck, 1986; Bretherton & Beeghly, 1982; Bukowski, 1990; Cain et al., 1997; Younger, Schwartzman, & Ledingham, 1986).

The vast majority of studies on the development of person perception to date have exposed children to a single piece of evidence about to-be-evaluated individuals. Though it is now well-established that showing children the same trait-relevant behavior multiple times improves their responding (e.g., Boseovski & Lee, 2006; Boseovski et al., 2013; Ferguson, Van Roozendaal, & Rule, 1986; but see Heller & Berndt, 1981), far less work has explored whether or not children can incorporate multiple *inconsistent* behaviors into their impressions. One study exploring this question demonstrated that children who were given one positive and one negative trait label about a target used an averaging strategy to generate liking judgments, whereby they liked inconsistent targets somewhat more than consistently bad targets and somewhat less than consistently good targets. These results were consistent across development from age 5, suggestive that even young children can incorporate inconsistency into their social evaluations (Hendrick et al., 1975). In contrast, other studies have demonstrated that young school-aged children have a relatively harder time reconciling inconsistencies in behavior than do older children, especially for traits not involving the prosocial-antisocial dimension (e.g., Mrug & Hoza, 2007; Rosenbach et al., 1973); these results suggest that abilities to reconcile inconsistencies in impression formation develop somewhat slowly across childhood. Indeed, to the extent that young children appear to evaluate characters who have behaved inconsistently, their evaluations are susceptible to various biases. These biases include a "positivity bias," whereby children sometimes ignore or discount negative behaviors during impression formation (e.g., Droege & Stipek, 1993; Lockhart, Chang, & Story, 2002; Mrug & Hoza, 2007; Rholes & Ruble, 1986), a "negativity bias," whereby children sometimes ignore or discount positive behaviors relative to negative behaviors (e.g., Aloise, 1993), as well as "recency effects," whereby young children's memory limitations lead them to focus on the trait-relevant behaviors they have observed most recently (e.g., Austin, Ruble, & Trabasso, 1977). The existence of these biases suggests that young children's ability to incorporate inconsistency into their impressions of social others is rather limited.

3. Can infants evaluate individuals who behave inconsistently?

To our knowledge, no research has yet tested whether infants' social evaluation system can track behavioral inconsistency. That said, a number of pieces of evidence are suggestive that infants might be able to do so, at least by 8–10 months of age. First, as reviewed above, by 8–10 months infants' social evaluations reflect fairly complex processing, such as identifying mental states and interpreting the same prosocial and antisocial behaviors differently across contexts (Hamlin, 2013b, 2014; Hamlin et al., 2011, 2013). Given the complexity of the judgments infants must make to succeed in these tasks, it seems reasonable to assume they can also aggregate across inconsistent behaviors. In addition, large bodies of evidence demonstrate that even very young infants are adept statistical learners, readily extracting patterns of information from their environment and using these patterns to inform inductive generalization (Kirkham, Slemmer, & Johnson, 2002; Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Xu & Garcia, 2008). Although the majority of statistical learning work to date has explored how infants extract patterns from non-social stimuli, a recent study by Tummeltshammer, Wu, Sobel, and Kirkham (2014) demonstrated that 8-month-olds are sensitive to the consistency with which social agents signal the locations of hidden objects, suggestive that infants have some sensitivity to patterns of information in social domains. If infants can compute the statistical frequency with which various kinds of stimuli occur in their environment, and

use this information to inform their inductive inferences, they may also be able to compute the statistical frequency with which prosocial and antisocial events are performed by an agent, and use those frequencies to inform their social evaluations.

On the other hand, other work suggests that the ability to process socially inconsistent input likely emerges after infancy. For example, whereas very young children prefer to learn new object labels from labelers who were previously 100% accurate versus at least somewhat inaccurate, it is not until age 4 that children reliably learn from a labeler who was not 100% accurate, but nevertheless relatively more accurate than another labeler (Pasquini, Corriveau, Koenig, & Harris, 2007). Relatedly, although infants and toddlers readily infer the goals and preferences of individuals who consistently choose the same object (Luo & Baillargeon, 2005; Kushnir, Xu, & Wellman, 2010; Woodward, 1998), it may not be until age 4 that children infer that someone prefers an object that she chooses most, but not all, of the time (Hu, Lucas, Griffiths, & Xu, 2015). If one assumes that evaluating characters who are inconsistently prosocial and antisocial requires the same cognitive mechanisms as does evaluating agents who are inconsistently accurate or who show inconsistent object preferences, then infants may not be able to evaluate inconsistently prosocial or antisocial others.

4. The current studies

To investigate whether infants can form social evaluations of characters who behave inconsistently, we presented infants with modified version of the "box scenario" puppet show utilized in previous research (e.g., Hamlin & Wynn, 2011). We chose to study 9-month-olds as this age has reliably demonstrated capacities for evaluating individuals who are consistently prosocial and antisocial (Hamlin, 2015; Hamlin & Wynn, 2011; Hamlin et al., 2007) as well as capacities to engage in mentalistic and context-specific social evaluations (Hamlin, 2013b; Hamlin et al., 2011, 2013). In the box scenario, infants are shown a series of puppet interactions in which a Protagonist puppet demonstrates an unfulfilled goal of opening a box that contains an attractive object, while two other puppets look on. On alternating events, infants see one of the two puppets (the "Helper") help the Protagonist open the box, allowing him to achieve his goal; whereas the other puppet (the "Hinderer") prevents the Protagonist from opening the box, blocking his goal. Hamlin and Wynn (2011) demonstrated that following habituation to the box scenario, both 5- and 9-month-olds are significantly more likely to reach for the Helper over the Hinderer (see replication in Hamlin, 2013b, and failure to replicate by Salvadori et al., 2015). In contrast, in the present research, rather than each puppet *either* helping *or* hindering throughout the course of the experiment, one of the two puppets acts inconsistently: On some events he helps the Protagonist, and on other events he hinders the Protagonist. In order to maximize the possibility that infants would be able to distinguish the characters, we chose to make only one puppet behave inconsistently: the second puppet either always helps or always hinders the Protagonist, as in previous work.

If infants can incorporate behavioral inconsistency into their social evaluations, they should prefer a relatively more prosocial agent to a relatively more antisocial agent, and so their puppet choices should differ depending on whether the consistent character is consistently Prosocial or consistently Antisocial. Specifically, when the Consistent Character always helps the Protagonist, infants should prefer the Consistent Character to the Inconsistent Character because the Inconsistent Character sometimes hinders and so is relatively less prosocial. On the other hand, when the Consistent Character always hinders the Protagonist, infants should prefer the Inconsistent Character, because he sometimes helps and so is relatively more prosocial. In contrast, if infants cannot incorporate behavioral inconsistency into their social evaluations, infants may choose randomly between the Consistent and Inconsistent Characters, regardless of the relative valence of their previous social acts.

5. Experiment 1

5.1. Methods

5.1.1. Participants

All participants were full-term, typically developing infants between 8 months, 11 days and 9 months, 16 days. In Experiment 1, there were 32 infants (14 females; average age = 8;27); 16 infants were randomly assigned to the *Consistently Prosocial Condition*, and 16 to the *Consistently Antisocial Condition* (details below). An additional nine infants began the experiment, but were not included in the final sample due to fussiness (5 infants), failure to choose a puppet (2), procedural error (1), or parental interference (1).

5.1.2. Procedures

The Behavioural Research Ethics Board of the University of British Columbia approved all procedures. Infants viewed puppet events from their parent's lap at the end of a long table surrounded on three sides by black curtains; a fourth curtain could be lowered to occlude the puppet stage between events. A puppeteer performed events by placing his hands underneath the rear curtain; he wore a black shirt to cover his arms and was thus entirely hidden to infants. Parents were instructed to sit quietly with their infants and to not attempt to influence their attention in any way; first time visitors to the lab were given pre-instructions for positioning during the choice procedure (detailed below). Throughout the study, infants' attention was recorded online by a coder who watched the infant through a live video feed in another room; coders were blind to what the infant was watching.

5.1.3. Prosocial and antisocial events

All events began identically: A curtain rose to reveal a clear box containing an orange octopus toy; two monkeys, one wearing a green shirt and one wearing a blue shirt, sat at the rear corners of the stage, angled slightly inward toward the box.

At the start of each trial, the Protagonist (a gray elephant) emerged from behind the curtain at the center of the rear of the stage, and moved to one side of the box. He leaned down and toward the box twice, as though "looking" at the toy inside. The Protagonist then jumped up and onto the front corner of the box lid, and began a series of failed attempts to open it. On his first and second attempts, the Protagonist lifted and shook the lid to indicate a struggle, dropped it shut, and sat up facing the infants. During attempts 3 and 4, the Protagonist lifted and shook the lid and lowered it back down, all while continuously holding onto it with his head down. On the Protagonist's 5th and final attempt, a monkey (wearing either a blue or green shirt) resting at the opposite rear-corner of the stage intervened. The following describes what happened at this point for each event type.

During Prosocial events, the acting monkey moved straight forward, paused briefly next to the box, and then jumped up and grasped the corner of the box lid opposite to where the Protagonist was currently struggling. The monkey then opened the box together with the Protagonist via a smooth, controlled motion. Once the box lid was open, the Protagonist jumped to a sitting position on top of the open lid and then laid his head down inside the box, grasping the toy inside. The monkey then jumped forward to the stage next to the box (so that infants would look toward

him), and then ran off-stage, slightly shaking as he moved backward to highlight his movements. Once the acting monkey was off-stage, all action paused.

During Antisocial events, the acting monkey moved forward, paused briefly next to the box, and jumped onto the corner of the lid opposite to where the Protagonist was currently struggling, slamming the lid shut. The Protagonist then jumped off the closed box, paused briefly, and then laid his head down next to the box without the toy. The acting monkey then jumped sideways to the stage next to the box and ran off-stage. As in Prosocial events, once the acting monkey was off-stage, all action paused.

During all events, the puppeteer watched the infant's face on a screen, and so was able to see if the infant looked away from the display. When this happened, the puppeteer simply paused the live show and waited for the infant to look back toward the display; the puppet show resumed from the point it left off as soon as the infant did so. This allowed us to ensure that all infants in the study saw the information conveyed in each event.

At the end of each event, attention coding began as soon as the acting monkey left the stage, as signalled to the coder in the control room by an audible mouth "click" from the puppeteer. Infants' attention toward and away from the paused scene was recorded from this point by a coder using the coding program *jHab* (Casstevens, 2007), continuing until infants looked away for two consecutive seconds or 30 s elapsed. Looking time was assessed in this way in order to determine when to present infants with the next trial.

Although previous work has often utilized a habituation procedure to examine social evaluation in infants of this age (Hamlin & Wynn, 2011; but see Hamlin et al., 2011), in order to effectively control the amount of information infants received about each character we chose not to habituate infants in the current studies. Instead, all infants viewed exactly 6 total trials, in which a "Consistent Character" performed 3 events and an "Inconsistent Character" performed 3 events. Consistent and Inconsistent Characters acted in alternation. The Consistent Character performed the same action on all 3 events (e.g., either always helping open the box, or always slamming the box closed), whereas the Inconsistent Character performed both kinds of actions across events (e.g., sometimes helping open the box, and sometimes slamming it closed). Half of the infants were randomly assigned to the *Consistently Prosocial Condition*, wherein the Consistent Character always helped the Protagonist open the box; half were randomly assigned to the *Consistently Antisocial Condition*, wherein the Consistent Character always prevented the Protagonist from opening the box.

Within each Consistency Condition, we also varied the extent of the variation between the two characters' behaviors via two different contrast conditions. In the *Easy Contrast Condition*, the Inconsistent Character more often acted differently from the Consistent Character, performing a different act on 2 of his 3 trials. That is, when the Consistent Character acted prosocially during all 3 events, the Inconsistent Character was antisocial twice and prosocial once. When the Consistent Character acted antisocially during all 3 events, the Inconsistent Character was prosocial twice and antisocial once. In the *Difficult Contrast Condition*, the Inconsistent Character more often acted similarly to the Consistent Character, performing a different act on only 1 of his 3 trials. That is, when the Consistent Character was prosocial, the Inconsistent Character was prosocial twice and antisocial once; when the Consistent Character was antisocial, the Inconsistent Character was antisocial twice and prosocial once. We reasoned that infants might be better able to distinguish the Consistent from the Inconsistent Characters in the Easy Contrast Condition, in which the characters' actions were most different.

Finally, in examining whether infants could incorporate behavioral variability into their social evaluations, we wanted to ensure that infants had sufficient opportunity to recognize that the two characters (sometimes) behaved differently. Therefore, during the first pair of events, the Consistent and Inconsistent Characters always performed different actions: If the Consistent Character was prosocial, the Inconsistent Character was antisocial and vice versa. During the Easy Contrast Condition the Inconsistent Character behaved similarly to the Consistent Character on either his second or third event (counterbalanced); therefore, half the babies in the Easy Contrast Condition saw both characters perform different actions during both the first and second pair of events; the other half saw them perform different actions in the first and the last pair. All babies in the Difficult Contrast Condition saw the puppets perform different actions solely during the first pair.

5.1.4. Choice

Following the 6 familiarization events, parents were instructed to rotate their chairs clockwise 90° from the stage and to place their toes on a pre-set line on the floor. Parents were instructed to position infants facing straight ahead at the front of their laps with their child's rump just on top of the knees, holding them firmly around the lower abdomen to ensure good trunk support. Parents were instructed to close their eyes so as not to influence their child's choice. A research assistant (who had not seen the puppet events and so was unaware of their identities) came out from behind the curtain and kneeled down approximately 90 cm in front of the line, holding the puppets behind his or her back. He or she then established eye contact with the infant by saying "Hi!" (several times if necessary) and then "Look!" holding up the Consistent and Inconsistent Characters in front of the infants, approximately 30 cm apart and initially out of their reach (approximately 45 cm away; calibrated by the choice presenter based on the size of the infant). Infants were required to look at each puppet and then back to the presenter; infants who did not do so quickly on their own were directed to do so by the presenter lightly shaking the puppets, at first together, and then one at a time if necessary. If the presenter could not re-establish eye contact with the infant following their looks toward each puppet, s/he repeated "Hi!" until eye contact was established. Once the infant was looking toward the presenter, the puppets were moved forward to within the infant's reach, approximately 30 cm away (again, depending on the size of the infant). Choice was determined online by this presenter, as the first puppet infants touched via a visually guided reach (i.e., a touch to a puppet preceded by a look to that puppet). If the presenter was not positive what the first visually guided reach had been by the end of the procedure, s/he (or another research assistant, also blind to the puppets' identities) determined the choice from videotape. Infants who did not make a choice within 2 min from the time the puppets were moved within reach were identified as having failed to make a choice and were excluded from the final sample. This two-minute threshold was determined prior to data collection and applies to all studies in our lab. For a measure of choice reliability, a second coder (who was blind to the identity of the puppets and to the choice decision of the presenter) later watched videos recordings and made choice evaluations for a randomly selected 25% of the subjects in each experiment.

5.1.5. Counterbalancing

The following were counterbalanced for infants in each condition: (1) shirt color of the Consistent Character (green/blue); (2) Consistent Character side during familiarization (left or right of stage); (3) Consistent Character order (first or second); (4) Consistent Character side during choice (left or right); and (in Easy Contrast Conditions) (5) the order in which the Inconsistent Character switches from its initial behavior to the oppositely valenced behavior (his second act or third act).

5.2. Results

The agreement on infants' choices between the choice presenters and the reliability coder was 100%. Patterns of choice did not differ significantly across any comparison. Infants in the *Consistently Prosocial Conditions* did not prefer the Consistently Prosocial Agent (Prosocial 3×) over the Somewhat Antisocial Agent (Prosocial 2×, Antisocial 1× (Hard Contrast), or Prosocial 1×, Antisocial 2× (Easy Contrast)): 8 infants chose the Consistently Prosocial Agent and 8 infants chose the Somewhat Antisocial Agent (binomial test, $p = 1.0$). Infants in the *Consistently Antisocial Conditions* did not prefer the Somewhat Prosocial Agent (Antisocial 2×, Prosocial 1× (Hard Contrast), or Antisocial 1×, Prosocial 2× (Easy Contrast)) over the Consistently Antisocial Agent (Antisocial 3×): 9 infants chose the Somewhat Prosocial Agent and 7 infants chose the Consistently Antisocial Agent (binomial test, $p = 0.80$). There were no differences in patterns of response between infants in the Easy and Hard Contrast Conditions (Fisher's Exact Test, $p = 1$), nor were there effects of order of events, order of inconsistent action, color of puppet, or side of puppet for any comparison (all $ps > 0.15$). Thus, infants' choices suggest that they cannot evaluate characters who perform inconsistent actions.

5.3. Discussion

When shown two agents, one of whom was relatively more prosocial than the other, infants failed to prefer the nicer agent. This was the case whether the Inconsistent Character was more versus less distinguishable from the Consistent Character (in Easy versus Difficult Contrasts, respectively), and did not depend on factors including action order, shirt color, or side. These results suggest that at 9 months of age, infants may have difficulty incorporating behavioral inconsistency into their social evaluations.

Notably, Experiment 1, particularly the Easy Contrast Condition, was designed to be relatively easy for infants: Our intention was to establish a baseline pattern of responding to comparisons involving minimal inconsistencies, so that we might compare data from future experiments to that baseline. As this was our goal, the design of Experiment 1 allowed for several alternative strategies that infants *might* have utilized to distinguish between the puppets; these strategies would have led to systematic responding during the choice procedure without reflecting an ability to evaluate characters who behave inconsistently. First, because Consistent and Inconsistent Characters always acted differently on their very first events, infants could have utilized a primacy bias, often referred to as a first-impression bias, to distinguish them. The first-impression bias refers to the tendency for the information one receives first about a target to contribute more strongly to one's impression of that target than does information received later; this bias has been demonstrated to influence adults' person perception (e.g., Anderson, 1965; Anderson & Hubert, 1963). Suggestive that infants could have generated positive and negative impressions of the characters based solely on their first (distinct) social acts, past work has demonstrated that 8-month-olds evaluate helpers and hinderers in the box scenario after just one instance of helping and one of hindering (specifically, infants' evaluations are sufficient for them to generate positive evaluations of those who subsequently help the former helper and hinder the former hinderer; Hamlin et al., 2011). However, whatever first impressions infants in Experiment 1 may have generated from the first pair of trials, they were clearly not sufficient to allow infants to prefer the more prosocial character during choice. These results suggest that seeing inconsistent information may be sufficient to wipe out previously established social evaluations.

Infants in the Easy Contrast Conditions might also have used a recency bias to differentiate the more prosocial from less prosocial characters in Experiment 1. Specifically, for half of the infants in the Easy Conditions the last pair of events included one prosocial and one antisocial event, which (as outlined in the previous paragraph) past work suggests should have been sufficient for infants at this age to generate social evaluations in the box scenario (Hamlin et al., 2011). Yet there was no evidence that infants who saw the characters perform different acts in the last pair of events were more likely to distinguish them than were infants who saw the Characters perform the same acts in the last pair (Fisher's Exact Test, $p = 1$). These results suggest that in social evaluation tasks, having seen inconsistent social information in the past is sufficient to disrupt the processing of more recent information.

A third strategy infants might have utilized in Experiment 1 would be to simply *ignore* the Inconsistent Character altogether, preferentially approaching the Consistent Character in the Consistently Prosocial Conditions and preferentially avoiding the Consistent Character in the Consistently Antisocial Conditions. That is, if infants found inconsistency too complicated to process, they might have stopped "trying" to attribute any sociomoral significance to the inconsistent agents' behavior at all, effectively rendering the Inconsistent Characters neutral. Although it has not been previously demonstrated that infants at this age prefer Helpers to Neutral characters and Neutral characters to Hinderers in the box scenario in particular, this has been demonstrated in another commonly utilized scenario, helping and hindering a Protagonist's goal to climb a hill (Hamlin et al., 2007), suggestive that infants might have used the same strategy here. As with the possibilities of utilizing first and recent impressions, that infants did not simply ignore the Inconsistent Characters in Experiment 1 suggests that there is something about behavioral inconsistency that disrupts social evaluation in 9-month-old infants.

Finally, results from Experiment 1 provided no evidence that infants utilize valence-based biases in their evaluations of characters who behave inconsistently. Studies with children have demonstrated that they sometimes discount negative information about individuals during impression formation, generally thought to reflect a "positivity bias" (e.g., Boseovski, Shallwani, & Lee, 2009; Droege & Stipek, 1993; Lockhart et al., 2002; Mrug & Hoza, 2007; Rholes & Ruble, 1986). If infants in Experiment 1 did the same, they might have failed to distinguish the characters in the Consistently Prosocial Condition (in which ignoring negative information would have made the characters more similar) but should have distinguished them in the Consistently Antisocial Condition (in which ignoring negative information would maintain the Inconsistent Character's relative prosociality). In contrast to positivity biases, other work has demonstrated the presence of a negativity bias in both social and non-social domains in development (see Vaish, Grossmann, & Woodward, 2008, for review). This has been demonstrated in social evaluation in infancy in particular, whereby infants appear to demonstrate relatively stronger dislike for antisocial others than like for prosocial others (e.g., Hamlin & Wynn, 2012; Hamlin et al., 2010). If infants in Experiment 1 were susceptible to a negativity bias, they might have failed to distinguish characters in the Consistently Antisocial Condition (where everyone is somewhat negative) but should have distinguished them in the Consistently Prosocial Condition (where one character never performs a negative act). In sum, we observed no evidence that either positivity or negativity biases influenced infants' responding in Experiment 1.

Results from Experiment 1 are the first to suggest that infants cannot incorporate behavioral inconsistency into their social evaluations. However, one alternative explanation for the null results observed is that the preference for helpers over hinderers in the box scenario is just not particularly robust, and may disappear

given small differences in the experimental procedure or subject population (see, e.g., Salvadori et al., 2015). For example, although there was a single puppeteer for all participants in Experiment 1 (C. Steckler), who was trained by Hamlin, he may have inadvertently performed box events slightly differently from those presented to infants in Hamlin et al.'s past published research, and these differences may have disrupted the tendency for infants to make social evaluations at all. Alternatively or in addition, the puppets used in Experiment 1 (large monkeys) had to our knowledge not previously been used in any social evaluation tasks. It is possible that infants disliked these puppets or were somehow distracted by them, and so were less able or willing to evaluate them based on their social acts in Experiment 1 than infants in past research. Finally, although as mentioned above past work has suggested that just one instance of helping and hindering behavior in the box show is sufficient to influence infants' evaluations of characters who subsequently help and hinder these characters (Hamlin et al., 2011), to date no study has specifically examined infants' basic preference for helpers over hinderers in the box scenario without utilizing a full habituation paradigm. Thus, perhaps infants failed to distinguish characters in Experiment 1 because our 6-event familiarization paradigm did not provide infants with sufficient exposure to the events.

To address these possibilities, in Experiment 2 we sought to replicate infants' preference for consistently prosocial over consistently antisocial characters in the box show, using the same experimenter, the same puppets, and the same familiarization method as in Experiment 1. If infants fail to prefer the Prosocial Character in Experiment 2, it would suggest that something outside of difficulty with behavioral inconsistency accounts for the null results in Experiment 1. On the other hand, if Experiment 2 replicates past work demonstrating a preference for consistently prosocial over consistently antisocial others, it would suggest that it is behavioral inconsistency in particular that disrupted social evaluation in Experiment 1.

6. Experiment 2

6.1. Methods

6.1.1. Participants

In Experiment 2, there were 16 infants (11 females; average age = 8 months, 27 days; range = 8;11–9;15). An additional six infants began the experiment, but were not included in the final sample due to failure to choose a puppet within 2 min (5), and procedural error (1).

6.1.2. Procedures

All procedures and stimuli were identical to that of Experiment 1, except that all participants viewed a Consistently Antisocial Character and a Consistently Prosocial Character act in alternation for six trials. After watching these events, infants chose between the Consistently Prosocial and Consistently Antisocial Characters.

6.1.3. Counterbalancing

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Prosocial Character (green or blue); (2) Prosocial Character side during familiarization (left or right of stage); (3) Prosocial Character order (first or second); (4) Prosocial Character side during choice (left or right).

6.2. Results

The agreement on infants' choices between the choice presenters and the reliability coder was 100%. Fifteen out of 16 infants preferred the Consistently Prosocial Character over the Consistently Antisocial Character (94%; binomial test, $p = 0.001$). There was no effect of order of events, color of puppet, or side of puppet during familiarization or choice on infants' choices.

6.3. Discussion

Infants' preference for the Consistently Prosocial over the Consistently Antisocial Character in Experiment 2 replicates past work demonstrating that infants prefer prosocial to antisocial agents in the box scenario, and suggests that it was not the puppeteer, the familiarization method, or the stimuli that was responsible for the documented null effects in Experiment 1. Together, results from Experiments 1 and 2 support the possibility that 9-month-old infants are unable to incorporate behavioral inconsistency into their social evaluations.

What kind of constraints or difficulties might account for infants' failure in Experiment 1? One possibility is that infants have trouble with *any* kind of inconsistent behavior performed by an agent, whether or not the inconsistency involves morally relevant behaviors. For instance, in Experiment 1, we demonstrated behavioral inconsistency by having the Inconsistent Character perform one type of goal-oriented action (e.g., opening a box with an agent) as well as its functional opposite (e.g., closing a box on that same agent). Previous research has found that infants at this age readily attribute goals to agents who repeatedly pursue the same goal (e.g., Woodward, 1998); however, less is known about what infants make of agents who exhibit two opposing goals. That is, perhaps infants' failure in Experiment 1 was not due to an inability to generate graded social evaluations about characters who are sometimes nice and sometimes mean, but due to difficulty interpreting any situation in which an agent first performs a goal, and then its opposite.

A related but somewhat higher-level explanation concerns how seeing opposite events might have influenced infants' sociomoral interpretation of the prosociality/antisociality of the events themselves. For instance, infants may have initially formed a sociomoral evaluation of the Inconsistent Character based on his first act (for example, liking him if he demonstrated a prosocial goal to open the box). But then during his subsequent inconsistent act(s), infants viewed him perform exactly the opposite behavior as what he had done before. This might have lead infants to (implicitly) reinterpret what box opening really signifies, calling into question their initial interpretation of the sociomoral significance of the event itself. That is, upon seeing the Inconsistent Character's inconsistent act, infants may have "thought" something akin to "if a character can both open and close the box in this case, it must not be prosocial or antisocial to do so." Because Consistent Characters performed one or the other of the same events that infants were lead to reinterpret, this disruption may have also influenced infants' evaluation of Consistent Characters, despite the fact that they always performed the same act.

Together, these explanations for infants' failures raise the (rather ironic) possibility that, in our first attempt to explore whether infants can evaluate individuals whose actions reflect the behavioral variation observable in the real world, we may have inadvertently set up an implausible situation: Individuals do not tend to do one thing to one individual and then the opposite thing to the same individual. Rather, when individuals behave inconsistently it tends to be in different situations, and typically toward different targets; for example, a sibling may be nice to most people, except to her little sister whom she always treats terribly; or, a mobster might be generally antisocial but treat his mother extremely well. Because in our previous experiments we varied *what type* of behavior the Inconsistent Character performed, but did not vary *whom* the behavior was directed toward; we failed

to consider that much real-world sociomoral behavioral variation stems from individuals interacting differently with different targets. Consistent with the possibility that infants found the inconsistent interactions in Experiment 1 to be implausible, a recent study suggests that infants form representations of social interactions involving two agents and expect those relationships to be consistent: After being familiarized to a Giver who gives apples and a Taker who takes apples, 12-month-olds looked longer when the Giver subsequently took from the target he previously gave to, and longer when the Taker subsequently gave to the target he previously took from (Tatone, Geraci, & Csibra, 2015).

One additional possibility is that infants *can* distinguish characters who behave inconsistently, but that such preferences are simply weaker than infants' preferences for characters who behave consistently. Indeed, the 'evaluative distance' between characters who are always nice versus always mean is likely wider than the distance between characters who are (for example) always nice versus sometimes mean; perhaps this distance is too small to demonstrate significant differences at the current sample sizes. That is, Experiment 1 utilized the same number of subjects per condition as in previous work comparing consistent versus consistent characters (e.g., Hamlin & Wynn, 2011; Hamlin et al., 2007; note that previous work using smaller sample sizes to Experiment 1 has shown that infants can distinguish consistent from neutral characters, but this has not been demonstrated using the current scenario), and so may not have been sufficiently powered to detect differential preferences for consistent versus inconsistent characters.

Experiment 3 was designed to address the two possibilities raised above for why infants failed to distinguish characters in Experiment 1. First, Experiment 3 explored the possibility that infants' failure stemmed from the implausibility of observing someone treat the very same individual in opposite ways, rather than a general inability to evaluate characters who behave inconsistently. To address this, in Experiment 3 we added an additional Protagonist, so that the Inconsistent Character could direct its inconsistent act toward someone new. We reasoned that this would provide a plausible reason for why the Inconsistent Character's behavior had changed (e.g., perhaps the Inconsistent Character feels differently about Protagonist B than Protagonist A; perhaps he treats some individuals well and some individuals poorly). Specifically, we switched to the second Protagonist during the pair of trials when the Inconsistent Character behaved inconsistently. The Consistent Character also acted on the new Protagonist during the same event pair; however, he performed the same act he had during previous events. Second, in order to address the possibility that the effect size for infants' ability to distinguish inconsistent sociomoral agents in the box show is non-zero, but nonetheless smaller than that of distinguishing consistent sociomoral agents, we doubled the sample size from Experiments 1 and 2, for a total of 32 infants per condition.

7. Experiment 3

7.1. Methods

7.1.1. Participants

Sixty-four infants participated in Experiment 3. Thirty-two infants were randomly assigned to the Consistently Prosocial Condition (19 females; average age = 9 months; 0 days; range = 8;12–9;18) and 32 infants to the Consistently Antisocial Condition (15 females; average age = 9 months; 2 days; range = 8;12–9;14). An additional 22 infants began or completed the procedure, but were not included in the final sample due to failure to choose a puppet (12 infants), fussiness (5), parental interference (2), or procedural error (2).[1]

7.1.2. Procedures

All procedures were identical to that of the Easy Condition of Experiment 1, except for the sample size increase and that the Protagonist changed from a gray elephant puppet to a white and brown dog puppet for the familiarization trial during which the Inconsistent Character switched its behavior from Prosocial to Antisocial or from Antisocial to Prosocial. In order to keep events between the Consistent and Inconsistent Characters as balanced as possible, the Protagonist was also the same New Protagonist dog puppet for one event involving the Consistent Character also; this New Protagonist event involving the Consistent Character was either immediately before or immediately the event during which the Inconsistent Character interacted with the New Protagonist dog (counterbalanced). All other trials involved the elephant Protagonist that was utilized in all other experiments. That is, in Experiment 3, 4 of the 6 familiarization events involved the elephant Protagonist (2 events in which the Consistent Character acted and 2 events in which the Inconsistent Character acted) The other 2 familiarization events involved the New Protagonist dog, which the Consistent and Inconsistent Characters either both helped (in the Consistently Prosocial Condition) or both hindered (in the Consistently Antisocial Condition).

7.1.3. Counterbalancing

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Consistent Character (green/blue); (2) Consistent Character side during familiarization (left/right of stage); (3) Consistent Character order (first/s); (4) Consistent Character side during choice (left/right); (5) the trial in which the Inconsistent Character acts on the dog (second act/third act); and (6) the trial in which the Consistent Character acts on the dog (immediately before/after the Inconsistent Character does).

7.2. Results

The agreement on infants' choices between the choice presenters and the reliability coder was 93.75%.[2] Patterns of choice did

[1] We began running Experiment 3 just after a period of lab construction, during which significant cosmetic changes were made in the lab. These included bright sky-blue walls, bright yellow doors, and curtains that were dark blue on the side walls but light blue at the back wall. These decisions were made for various reasons, but mostly to make the lab environment more fun and kid-friendly. The decisions, however, also appeared to make the lab less science-friendly: over the course of several months in this new lab space we noticed that infants seemed more distracted than before the construction, and that studies that had been showing significant effects before the construction (in two different lab spaces in the department) were now showing null effects across the board (typically in the form of persistent color preferences, especially if one of the puppets was wearing blue). The stark differences before and after construction, in data from other experiments not described in this manuscript, made us unsure why infants were demonstrating null effects in the conditions for this experiment, and so the lab decided to shelve all data collected in all puppet show studies in the lab since construction. The (shelved) results for Experiment 3 that was completed in the renovated lab space were, as mentioned, also not significant (Fisher's exact test, $p = 0.14$): In the Consistently Prosocial Condition, 6/13 infants chose the relatively nicer agent, binomial test $p = 1$; in the Consistently Antisocial Condition, 12/16 infants chose the relatively nicer agent, binomial test $p = 0.077$. We then repainted the lab's walls and doors off-white and went back to black curtains, in an attempt to reduce the distractibility of the environment. Although participants in this condition continued to demonstrate null effects after these changes, other studies returned to the patterns we had been seeing before construction, suggesting our suspicions that the renovated lab was too distracting were correct, despite that they did not account for this particular set of null results. No other conditions in the reported set of experiments included infants who completed the procedure in the problematic lab space.

[2] There was one discrepancy. To resolve this, another coder evaluated the infant's choice and was in agreement with the original presenter. We thus considered the infant's choice to be as the original presenter had decided.

not differ significantly across any comparison (Fisher's Exact Test, $p = 0.45$). Infants in the *Consistently Prosocial Condition* did not prefer the Consistently Prosocial Character (Prosocial 3×, once toward a new Protagonist) over the Mostly Antisocial Character (Antisocial 2×, Prosocial 1× toward a new Protagonist): 16 infants chose the Consistently Prosocial Character and 16 infants chose the Mostly Antisocial Character (binomial test, $p = 1.0$). Infants in the *Consistently Antisocial Condition* did not prefer the Mostly Prosocial Character (Prosocial 2×, Antisocial 1× toward a new Protagonist) over the Consistently Antisocial Character (Antisocial 3×, once toward a new Protagonist); 20 infants chose the Mostly Prosocial Character and 12 infants chose the Consistently Antisocial Character (binomial test, $p = 0.22$). There were no significant effects of the order of events, which pair contained the inconsistent act, the color of the puppets, or side for any comparison (all $ps > 0.07$).

7.3. Discussion

Neither doubling the sample size nor changing the identity of the Protagonist during the inconsistent event provided any evidence that infants distinguish between the two characters. That is, even by providing infants with a plausible reason for the Inconsistent Character's inconsistent behaviors (that he might hold different opinions about different targets), infants still failed to prefer the relatively more Prosocial agent during choice. This failure is consistent with past work with young children showing that presenting young children with an actor acting prosocially or antisocially toward different Protagonists does not facilitate trait attribution (Boseovski & Lee, 2006), and provides additional support for the possibility that behavioral inconsistency is particularly difficult for infants in the first year of life.

In Experiment 4, we sought to rule out one additional, domain-general possibility for why infants appear unable to incorporate behavioral inconsistency into their social evaluations. Specifically, it is possible that working memory difficulties might constrain infants' ability to update a previously established representation (of any kind) in light of new evidence. This possibility is reminiscent of previous research demonstrating that, although 11-month-old infants can update their representations of the quantity of each of two hidden arrays of objects when objects are added to one array at a time, they cannot update those representations when objects are added to the arrays in alternating order (Moher & Feigenson, 2013). Studies in this area have utilized both looking time/violation of expectation paradigms and "find the cracker" choice paradigms. In "find the cracker" studies, 11-month-olds watch as two quantities of crackers are hidden inside two opaque containers. After the hiding is finished, infants are allowed to choose which container to choose by crawling toward one or the other. It is assumed that if infants can remember how many crackers are in each container, they should crawl toward the one with more (Cheries, Mitroff, Wynn, & Scholl, 2008; Feigenson & Carey, 2003, 2005; Feigenson & Yamaguchi, 2009). Indeed, when all of the crackers are placed one-by-one in container A before any crackers are placed in container B, infants reliably crawl to whichever container has more (with some restrictions based on overall amount, see e.g., Feigenson & Carey, 2005). In contrast, when crackers are added to each container in alternating order, in that one is added to container A, then one to container B, then a second to A, etc., infants subsequently choose randomly between the two containers (Feigenson & Yamaguchi, 2009). These results suggest that although infants can hold two numerical representations in mind at once (for example, remembering how many crackers went into container A while they see crackers placed in container B), updating a previously established numerical representation while simultaneously holding another in mind and later returning to the first representation is overly taxing of their working memory, leading them to "drop" their initial representations and fail the task. Notably, control conditions confirmed that infants do not drop their initial representations simply because placing crackers in containers in an alternating order requires frequent attentional shifts: In a condition where infants see an object placed in A then have their attention drawn to B by an experimenter's waving hand (but no cracker placed in B and hence no representation update of B required), infants succeed when more crackers are placed in A (Feigenson & Yamaguchi, 2009 A(B)AB condition). Rather, it appears that it is updating one representation, and then needing to return to a previously established representation on a subsequent event, that is particularly difficult for infants at this age.

To apply this reasoning to infants' failures in the current Experiments 1 and 3, although past research suggests that infants can hold two social representations in mind at once (that one individual is prosocial and another is antisocial; Hamlin et al., 2011), and even maintain those representations over alternating events (e.g., Hamlin & Wynn, 2011; the current Experiment 2), perhaps infants are unable to *update* one social representation while simultaneously holding another previously established representation in mind, especially if they must return to the previously established representation once again. Therefore, we reasoned that presenting each character's actions in succession, rather than in alternating order, might ease the demand on infants' working memory, allowing them to successfully form social evaluations and prefer a relatively more prosocial character to a relatively less prosocial one, even when needing to update their representation of the Inconsistent Character when he behaves inconsistently.

8. Experiment 4

8.1. Methods

8.1.1. Participants

Sixty-four infants participated in Experiment 4. Thirty-two infants were randomly assigned to the Consistently Prosocial Condition (16 females; average age = 9 months, 1 day; range = 8;14–9;17) and 32 infants to the Consistently Antisocial Condition (16 females; Antisocial age = 9 months, 2 days; range = 8;12–9;17). An additional 14 infants began or completed the procedure, but were not included in the final sample due to failure to choose a puppet (3 infants), fussiness (6), procedural error (3), or parental interference (2).

8.1.2. Procedures

Given that Experiment 3 provided no evidence that changing the Protagonist facilitated infants' capacity to evaluate characters who behave inconsistently, and it is possible that adding a 4th character to the events places additional demands on infants' working memory, we used a single Protagonist in Experiment 4. All procedures were identical to that of the Easy Contrast Condition of Experiment 1, except that the Consistent Character and the Inconsistent Character did not act in alternation. Instead, each agent completed its acts consecutively; that is, the Consistent Character acted for three trials in a row before the Inconsistent Character acted for three trials in a row, or the reverse.

8.1.3. Counterbalancing

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Consistent Character (green/blue); (2) Consistent Character side during familiarization (left/right of stage); (3) Consistent Character order (first/s); (4) Consistent Character side during choice (left/right); and (5) the act in which the Inconsistent Character switches from its

initial behavior to the oppositely valenced behavior (his second act/third act).

8.2. Results

The agreement on infants' choices between the choice presenters and the reliability coder was 100%. As in previous Experiments, patterns of choice did not differ by valence of the consistent act (Fisher's Exact Test, $p = 1.00$). Infants in the Consistently Prosocial Condition did not prefer the Consistently Prosocial Character (Prosocial 3×) over the Mostly Antisocial Character (Prosocial 1×, Antisocial 2×); 18 infants chose the Consistently Prosocial Character and 14 infants chose the Mostly Antisocial Character (binomial test, $p = 0.60$). Similarly, infants in the Consistently Antisocial Condition did not prefer the Mostly Prosocial Character (Antisocial 1×, Prosocial 2×) over the Consistently Antisocial Character (Antisocial 3×); 18 infants chose the Mostly Prosocial Character and 14 infants chose the Consistently Antisocial Character (binomial test, $p = 0.60$). Turning to order effects, although most order effects were not significant (with $ps > 0.07$; consistent character start order, color of puppet, and puppet show side), there was one significant order effect. Specifically, for infants in the Consistently Prosocial Condition, there was a significant effect of order of inconsistent action, Fisher's Exact Test, $p = 0.001$. Breaking this down, infants who saw the Mostly Antisocial Character perform two antisocial actions and then a prosocial action were marginally *more* likely to select the Mostly Antisocial Character over the Consistently Prosocial Character (4/16 selected the Consistently Prosocial Character; binomial test, $p = 0.077$; this marginal effect runs counter to our hypothesis). By contrast, infants who saw the Mostly Antisocial Character perform one antisocial action, then one prosocial action, then a second antisocial action, were significantly more likely to select the Consistently Prosocial Character over the Mostly Antisocial Character (14/16 selected the Consistently Prosocial Character; binomial test, $p = 0.004$). This finding provides limited evidence that infants showed a moral recency effect, whereby their evaluations of agents were based on the last (most recent) piece of information they saw. However, because a moral recency effect did not show up in any other Experiment, nor in the Consistently Antisocial Condition within Experiment 4, the evidence that recency consistently influences infants' social evaluations is quite limited, and may reflect Type I error.

8.3. Discussion

Experiment 4 provided no evidence that presenting all of each agent's actions in a row, rather than in alternating sequence, improved infants' ability to form social evaluations of characters who demonstrated inconsistent behaviors – even despite the doubled sample size versus Experiment 1. This suggests that reducing demands on infants' working memory, at least in the manner attempted in this Experiment, is not sufficient to improve infants' responding, and further supports the claim that 9-month-old infants have particular difficulty incorporating behavioral inconsistency into their social evaluations.

9. General discussion

Across several attempts, we found that 9-month-old infants were unable to form social evaluations in the face of behaviorally inconsistent input; specifically, infants failed to prefer more-prosocial characters to more-antisocial ones when either character behaved inconsistently. In Experiment 1, infants failed to prefer a character who was always prosocial over one who was mostly (in the Easy Contrast Condition) or somewhat (in the Hard Contrast Condition) antisocial, and they failed to prefer a character who was mostly or somewhat prosocial over one who was always antisocial. Suggestive that this failure did not stem from minor differences in the stimuli or general methodology, infants in Experiment 2 preferred a character who was always prosocial over one who was always antisocial, replicating past demonstrations of infants' preference for prosocial others in the box scenario (Hamlin & Wynn, 2011) using a fixed number of trials. In two additional Experiments, various attempts to help infants distinguish between the two characters did little to increase systematic responding: In Experiment 3, the Inconsistent Character directed his behavior toward a different Protagonist, perhaps providing a plausible reason for why an agent would perform inconsistent acts; in Experiment 4 demands on working memory were reduced by having characters perform all of their actions in a row, based on previous work suggestive that this facilitates updating numerical representations (Feigenson & Yamaguchi, 2009). In both Experiments 3 and 4 sample sizes relative to Experiment 1 were doubled to account for the possibility that the effect size for infants' graded sociomoral preferences is smaller than their preferences between consistently prosocial and consistently antisocial individuals. Despite all this, in neither Experiment did infants reliably prefer the more prosocial character.

Notably, not only did infants fail to incorporate behavioral inconsistency into their social evaluations, they also failed to utilize various alternative (though biased) strategies to distinguish between characters. First, infants could have utilized a strategy whereby they ignored the inconsistent behavior performed by the Inconsistent Character, or ignored processing the sociomoral valence of the Inconsistent Character altogether (effectively making him a neutral character). That they did not do so suggests that inconsistent behavioral information may be difficult for infants to ignore, and yet difficult to incorporate into their social evaluations. Second, infants showed no evidence of valence-based biases, such as either downplaying negative information (as in a positivity bias; see, e.g., Boseovski et al., 2009; Droege & Stipek, 1993) or highlighting it (as in a negativity bias; see, e.g., Hamlin & Wynn, 2012; Hamlin et al., 2010; Vaish et al., 2008). Finally, aside from weak evidence for a recency effect in one condition of Experiment 4, infants did not show evidence of either primacy or recency biases.

If infants did not utilize any of the strategies listed above, and if their failures were not due to any of the possibilities tested in the current studies, then how are we to understand them? This question is particularly significant in light of past research whereby infants at this age form reliable social preferences of characters involved in complex sociomoral interactions, including contexts where mental states are inconsistent with outcomes (e.g., Hamlin, 2013b; Hamlin et al., 2011). In addition, the large infant statistical learning literature suggests that infants *are* capable of tracking the relative frequency of at least certain forms of information from quite early in development (e.g., Kirkham et al., 2002; Romberg & Saffran, 2010; Saffran et al., 1996, 1999; Xu & Garcia, 2008). Within the social domain specifically, a recent study found that while infants readily follow attentional cues from previously 100% reliable informants, they fail to learn from informants who provide informative information only 25% of the time (Tummeltshammer et al., 2014). Although there are various differences between this and other statistical learning papers and the current studies that could account for infants' failures, these results support the possibility that infants have particular difficulty with inconsistent information in the sociomoral domain.

Indeed, it is possible that infants' failure in the current studies reflects an exaggerated form of adults' difficulties with others' behavioral inconsistency. That is, although work with adults suggests that they are *able* to form aggregate representations of others' character over varied inputs, they may resist doing so (e.g., Heider,

1958). Instead, adults are willing to form stable social impressions from extremely limited behavioral information, ones that are highly resistant to subsequent updating in light of new information (e.g., Ambady & Rosenthal, 1992; Carlston & Skowronski, 1994; Winter & Uleman, 1984). This resistance may reflect some difficulty with behavioral inconsistency that persists into adulthood, which is especially salient early in development.

Of course, as this is the first demonstration that infants' have trouble with behavioral inconsistency when choosing between two agents, convergent evidence is clearly needed to support both the validity of the failures observed and to further elucidate their cause. This evidence should come from other experimental paradigms, other social and moral domains, and infants of other ages. In particular, although we ruled out some explanations for infants' failure to parse behavioral inconsistency in the current studies, we provided little evidence for what does explain the failure.

Future research should examine other ways to help infants succeed at these tasks. For instance, one reason infants failed our tasks might be because processing inconsistency requires relatively *more* data than does processing consistency. Perhaps if infants were habituated to an approximate proportion of inconsistency, rather than familiarized to an exact amount of inconsistency, they would succeed; similarly, future studies might include twice as many trials and just a small amount of inconsistency. A second possibility is that infants might succeed only if *both* characters act inconsistently, but to different degrees. Although at first glance this appears more difficult than what was tested in the current studies, it seems possible that infants are confused when one character acts consistently and the other acts inconsistently, leading them to focus more on the consistency information itself than on the sociomoral significance of the behaviors. A third possibility is that (despite changing the Protagonist in Experiment 3) our inconsistent behaviors were too "close" to each other, and infants would succeed if shown prosocial and antisocial behaviors across entirely different contexts, such as the hill (Hamlin et al., 2007), box, and ball scenarios (Hamlin & Wynn, 2011), with different Protagonists in each case.

Finally, we note that the current studies utilized only a single dependent variable – character choice – to explore infants' capacity for evaluating inconsistency. That infants failed here need not imply that they lack any ability to process inconsistency; in fact, infants' failure to evaluate inconsistent characters could be taken to imply that our participants were sensitive to the presence of inconsistency, but were nevertheless unable to incorporate it into their social evaluations. An alternative way to examine whether infants are sensitive to inconsistency might be to explore whether they find inconsistent behavior *surprising*: Infants might expect agents who have previously helped/hindered third parties to continue to help/hinder the same (and perhaps other) individuals in the future. Recent work suggests that by 12 months of age, infants hold such expectations for characters involved in giving and (permitted) taking interactions when subsequent acts are directed toward the same targets (Tatone et al., 2015); however, given the relatively small number of interactions our studies were not well-suited to explore whether infants found inconsistent behaviors surprising. Future work should examine to what extent infants expect helpers and hinderers to help and hinder in the future.

That infants do not appear to be able to incorporate behavioral inconsistency into their social evaluations has important implications for our understanding of the development of moral learning. Of course, moral learning is a multi-faceted issue, with many different possible learning mechanisms leading to many different possible learning outcomes. Here we note that moral learning is often cyclical: Just as young children use their evaluations of certain good and bad acts to determine *who* is good and bad (in infants, this might lead them to prefer certain individuals over others), they can also use their positive and negative evaluations of others to determine the value of the novel *actions* that others perform (in infants, this might lead them to preferentially perform certain acts and not others). Research to date has suggested that when prosocial and antisocial others behave consistently, even preverbal infants are capable of both types of moral learning (Hamlin & Wynn, 2012; Hamlin et al., 2007). However, given that results from the current studies suggest that even small amounts of inconsistency may disrupt infants' capacity to determine the relative value of prosocial and antisocial others in their environment, and in the 'real world' individuals rarely behave consistently 100% of the time, it may be that preverbal infants' capacities for learning who (and in turn what) is good and bad in the real world are fairly limited. The current studies suggest that one task of moral development, then, is to acquire or synthesize the ability to aggregate across inconsistent behaviors in order to form overall sociomoral evaluations of real world, variably behaving agents. A productive avenue for future work would be to examine when and how this development occurs.

Acknowledgements

We acknowledge the generous support of the Social Sciences Research Council of Canada, Grant #435-2014-2173, and thank the families who made this research possible.

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (1968). *Theories of cognitive dissonance: A sourcebook*. Chicago: Rand McNally.

Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attribution biases. *Journal of Experimental Child Psychology, 55*, 177–193.

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72*, 1409–1425.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology, 2*, 1–9.

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

Austin, V. D., Ruble, D. N., & Trabasso, T. (1977). Recall and order effects as factors in children's moral judgments. *Child Development, 48*, 470–474.

Benenson, J. F., & Dweck, C. S. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development, 57*, 1179–1187.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*, 1–29.

Boseovski, J. J., Chiu, K., & Marcovitch, S. (2013). Integration of behavioral frequency and intention information in young children's trait attributions. *Social Development, 22*, 38–57.

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–513.

Boseovski, J. J., Shallwani, S., & Lee, K. (2009). It's all good: Children's personality attributions after repeated success and failure in peer and computer interactions. *British Journal of Developmental Psychology, 27*, 783–797.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, 18*, 906–921.

Bukowski, W. M. (1990). Age differences in children's memory of information about aggressive, socially withdrawn, and prosociable boys and girls. *Child Development, 61*, 1326–1334.

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*, e88612. http://dx.doi.org/10.1371/journal.pone.0088612.

Burns, M. P., & Sommerville, J. A. (2014). "I pick you": The impact of fairness and race on infants' selection of social partners. *Frontiers in Psychology, 5*.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domain. *Social Development, 6*, 53–75.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology, 66*, 840–856.

Casstevens, R. M. (2007). *Jhab: Java habituation software (version 1.0.2) [computer software]*. Chevy Chase, MD: Java.

Cheries, E. W., Mitroff, S. R., Wynn, K., & Scholl, B. J. (2008). Cohesion as a constraint on object persistence in infancy. *Developmental Science, 11*, 427–432.

Choi, Y., & Luo, Y. (2015). 13-month-olds' understanding of social interactions. *Psychological Science, 26*, 274–283. http://dx.doi.org/10.1177/0956797614562452.

Dozier, M. (1991). Functional measurement assessment of young children's ability to predict future behavior. *Child Development, 62*, 1091–1099.

Droege, K. L., & Stipek, D. J. (1993). Children's use of dispositions to predict classmates' behavior. *Developmental Psychology, 29*, 646–654.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*, 568–584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*, 295–313.

Feigenson, L., & Yamaguchi, M. (2009). Limits on infants' ability to dynamically update object representations. *Infancy, 14*, 244–262.

Ferguson, T. J., Van Roozendaal, J., & Rule, B. G. (1986). Informational basis for children's impressions of others. *Developmental Psychology, 22*, 335–341.

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science, 14*, 1012–1020. http://dx.doi.org/10.1111/j.1467-7687.2011.01048.x.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.). *The innate mind* (Vol. 3, pp. 367–391). New York: Oxford.

Hamlin, J. K. (2013a). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science, 22*, 186–193. http://dx.doi.org/10.1177/0963721412470687.

Hamlin, J. K. (2013b). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition, 128*, 451–474. http://dx.doi.org/10.1016/j.cognition.2013.04.004.

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*, 1–10.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science, 16*, 209–226.

Hamlin, J. K. (2015). The case for social evaluation in preverbal infants: Gazing toward one's goal drives infants' preferences for Helpers over Hinderers in the hill paradigm. *Frontiers in Psychology, 5*, 1–9.

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

Hamlin, J. K., & Wynn, K. (2012). Who knows what's good to eat? Infants fail to match the food preferences of antisocial others. *Cognitive Development, 27*, 227–239.

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

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

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Science*. http://dx.doi.org/10.1073/pnas.1110306108.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heller, K. A., & Berndt, T. J. (1981). Developmental changes in the formation and organization of personality attributions. *Child Development, 52*, 683–691.

Hendrick, C. (1968). Averaging vs summation in impression formation. *Perceptual and Motor Skills, 27*, 1295–1302.

Hendrick, C., Franz, C. M., & Hoving, K. L. (1975). How do children form impressions of persons? They average. *Memory & Cognition, 3*, 325–328.

Hu, J., Lucas, C. G., Griffiths, T. L., & Xu, Fei (2015). Preschoolers' understanding of graded preferences. *Cognitive Development, 36*, 93–102.

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. *Advances in Experimental Social Psychology, 2*, 219–266.

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

Joyce, R. (2006). *The evolution of morality*. Cambridge: MIT Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In B. Thornton, R. M. Ryckman, & M. A. Robbins (1982). The relationships of observer characteristics to beliefs in the causal responsibility of victims of sexual assault. *Human Relations* 35(4), 321–330.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*, 1134–1140.

Lee, Y. E., Yun, J. E. E., Kim, E. Y., & Song, H. J. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PloS One, 10*(9), e0135588.

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's traitunderstanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development, 78*, 1543–1558.

Livesley, W. J., & Bromley, D. B. (1973). *Person perception in childhood and adolescence*. New York, London: J. Wiley.

Lockhart, K. L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of traits: Protective optimism? *Child Development, 73*, 1408–1430.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16*, 601–608.

Mrug, S., & Hoza, B. (2007). Impression formation and modifiability: Testing a theoretical model. *Merrill-Palmer Quarterly, 53*(4), 631–659.

Moher, M., & Feigenson, L. (2013). Factors influencing infants' ability to update object representations in memory. *Cognitive Development, 28*, 272–289.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216–1226.

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*, 550–560.

Rholes, W. S., & Ruble, D. N. (1986). Children's impressions of other persons: The effects of temporal separation of behavioral information. *Child Development, 54*, 872–878.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 906–914.

Rosenbach, D., Crockett, W. H., & Wapner, S. (1973). Developmental level, emotional involvement, and the resolution of inconsistency in impression formation. *Developmental Psychology, 8*, 120–130.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10*, 173–220.

Ruble, D. N., & Dweck, C. S. (1995). Self-conceptions, person conceptions, and their development. In N. Eisenberg (Ed.), *Review of personality and social psychology. Social development* (Vol. 15, pp. 109–139). Thousand Oaks, CA: Sage.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

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*, e0140570.

Scola, C., Holvoet, C., Arciszewski, T., & Picard, D. (2015). Further evidence for infants' preference for prosocial over antisocial behaviors. *Infancy, 20*, 684–692.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96*, 58–83.

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.

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science, 25*, 1730–1738.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383–403.

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*, 237–252.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, 105*, 5012–5015.

Younger, A. J., Schwartzman, A. E., & Ledingham, J. E. (1986). Age-related differences in children's perceptions of social deviance: Changes in behavior or in perspective? *Developmental Psychology, 22*, 531.

Yuill, N. (1992). Children's conception of personality traits. *Human Development, 35*, 265–279.