Failed attempts to help and harm: Intention versus outcome in preverbal infants’ social evaluations

J. Kiley Hamlin

The University of British Columbia, Canada

Abstract

Mature moral judgments include an analysis of both the outcomes of others’ actions as well as the mental states that drive them. While adults easily incorporate both intention and outcome into their moral evaluations, scores of developmental studies suggest that it may be uniquely difficult for young children to privilege intention in their judgments of right and wrong (e.g., Piaget, 1932/1965), leading to the conclusion that the ‘moral mind’ of the young child is fundamentally different from that of older children and adults. The current studies utilize a puppet-choice methodology shown to provoke reliable social preferences throughout the first year after birth (e.g., Hamlin, Wynn, & Bloom, 2007), and provide evidence that 8-month-old infants incorporate, and even privilege, intentions in their social evaluations. In contrast, 5-month-olds appear only able to distinguish characters who intend the outcomes they cause. Such results suggest that one requirement for mature moral judgments, the ability to distinguish between intentions and outcomes in morally relevant events, is present by 8 months of age.

1. Introduction

Notions of moral and criminal responsibility rely critically on the mental states that drive action; for instance, what an individual wanted to do, what his or her motives, beliefs, or intentions were, and how those related to what actually happened. Legal systems require that a guilty mind (mens rea) accompany a guilty act (actus reus) for a perpetrator to be held criminally responsible (e.g., Hart, 1968), and adults consider a perpetrator’s beliefs and intentions in addition to an action’s consequences when assigning praise and blame (e.g., Baird & Moses, 2001; Cushman, 2008; Killen & Smetana, 2008; Malle, 1999; Mikhail, 2007; Young, Cushman, Hauser, & Saxe, 2007). Indeed, when the link between one’s mental states and actions is severed completely (such as when acting under duress or with diminished mental capacity) criminal and moral responsibility may be entirely eliminated (e.g., Fincham & Roberts, 1985; Newman & Weitzer, 1956).

Of course, when distributing punishments, courts of law and lay individuals consider both intentions and outcomes: thinking bad thoughts without acting on them is rarely punished, attempting murder is not as punishable as actually murdering, and, depending on circumstance, involuntary manslaughter could be punishable despite lacking negative intent. Yet, given that some act was performed, judgments of how wrong the action was or how blameworthy the actor is tend to focus specifically on the actor’s mental states; the difference between someone or something being good versus morally good, and bad versus morally bad, often rests solely on a mentalistic distinction (Berndt & Berndt, 1975; Cushman, 2008; Kant, 1785/1959; Keasey, 1978; Malle, 1999). Thus, mentalistic moral reasoning in particular is critically related to human moral cognition in general.

In addition to playing a defining role in morality, it is presumably functional to focus on mental states during any social-cognitive task, morally relevant or not. Although
individuals’ immediate behaviors likely have much to do with constraints specific to the particular situations they are in, arguably at least some mental states persist through time and play a role in determining future behaviors (e.g., Gopnik & Wellman, 1994; Heider, 1958; Jones & Davis, 1965; Kelley, 1972; Malle, 1999; c.f. Doris, 2002). By tracking mental states, then, observers can gain some ground in the social-cognitive tasks of accurately predicting individuals’ future behaviors and reasoning about their novel actions in real time. Tracking others’ valenced mental states (that is, ones related to bringing about good and bad effects in the world such as helping and harming) might be particularly beneficial: despite the imperfect relationship between mental states and immediate behaviors, individuals who currently possess harmful intentions are presumably somewhat more likely to cause harm in the future than those who currently possess helpful intentions, over and above whatever outcomes they have currently brought about. Thus, systematically avoiding those who have previously demonstrated malevolent mental states might allow observers to escape personal harm, whereas systematically associating with those who have demonstrated benevolent mental states might lead to future personal benefit. Beginning with Darwin (1871), scholars have hypothesized that the capacity to distinguish ‘good guys’ from ‘bad guys’ was so beneficial to group-living that it was selected for over evolutionary time, and is itself the foundation of human morality (e.g., Alexander, 1987; Cosmides & Tooby, 1992; deWaal, 2006; Henrich & Henrich, 2007; Joyce, 2006; Katz, 2000; Price, Cosmides, & Tooby, 2002).

The defining and functional role of mental state analysis in humans’ social and moral cognitive capacities has led to a vast amount of research into their developmental origins: When does mentalizing, particularly within the moral domain, emerge in childhood, and what processes account for its emergence? Beginning with the seminal work of Jean Piaget in The Moral Judgment of the Child (1932/1965), scores of developmental studies suggest that mentalistic moral evaluation is a relatively late developmental achievement, emerging sometime during middle childhood (around or after 7 years). Piaget asked children of various ages which of two child perpetrators was naughtier: John, who broke 15 cups as a side effect of a positively-intended behavior, or Henry, who broke 1 cup as a side effect of a negatively-intended behavior. Specifically, children were asked to judge relative naughtiness when the valence of one actor’s intention was pitted against the valence of the outcome experienced by another.Piaget reported that children under 7–10 years tended to evaluate naughtiness based on outcome (John naughtier than Henry), whereas older children focused on intent (Henry naughtier than John). Through evidence from this and several other related comparisons, Piaget concluded that children are initially “moral realists” who assign explicit blame based on objective states of the world like outcomes, rather than on subjective states of the mind like intentions.

Piaget’s contribution to the study of moral development is unparalleled, and his views remain extremely influential (see, e.g., Carpendale, 2009; Lapsley, 1996; Müller, Carpendale, & Smith, 2009). Yet, in the decades since The Moral Judgment of the Child (1932), the conclusion that children do not consider mental states in their moral judgments until well into middle childhood has been challenged, most commonly on methodological grounds (see Karniol, 1978, for review). In particular, Piaget’s tasks require fairly advanced cognitive skills, such as processing complicated verbally-presented stories, integrating several pieces of conflicting information into the same judgment, and explicitly verbalizing one’s knowledge. It is possible that limitations in these domain-general capacities led Piaget to underestimate young children’s moral abilities. Supporting these challenges, many studies suggest that if Piaget’s methodologies are simplified in various ways such that stories are presented visually, intention is de-confounded from outcome, motives are made explicit rather than merely implied, and outcomes result directly from intended acts rather than as side-effects, children consider mental states years earlier than Piaget reported, in some studies as early as age 3 (e.g., Chandler, Greenspan, & Barenboim, 1973; Farnhill, 1974; King, 1971; Nelson, 1980; Yuill, 1984; Zelazo, Helwig, & Lau, 1996 see Karniol, 1978; Keasey, 1978, for reviews).

Yet reducing the age at which children begin to consider mental states does not itself refute Piaget’s claims for developmental change: determining exactly when a change occurs in development is arguably less meaningful than demonstrating that it occurs at all (e.g., Carpendale, 2009; Piaget, 1932/1965). Since Piaget’s initial work there have been literally hundreds of studies of the intention/outcome in children; these have consistently demonstrated that young children give more weight to outcomes than older children do, and initially privilege outcome over intention if they conflict (e.g., Baird & Astington, 2004; Chandler, Sokol, & Hallett, 2001; Costanzo et al., 1973; Farnill, 1974; Hebble, 1971; Imamoglu, 1975; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Leon, 1980; Zelazo et al., 1996). Thus, results demonstrating earlier competence moderate, but also essentially support, Piaget’s claims.

But perhaps there are more significant reasons to revisit Piaget’s conclusions. First, although moral judgments are clearly modified and adjusted throughout the lifespan (indeed, well beyond the notion that mental states are more important than outcomes (see Sokol, Chandler, & Jones, 2004)), it is well documented that young children engage in some morality-specific judgments in early childhood. For instance, whereas Piaget contended that children’s earliest moral judgments focus solely on whether a rule was broken and/or if a perpetrator would be punished, “domain theorists” have convincingly demonstrated that even very young children distinguish between different kinds of rule-violations, viewing only rules within the moral domain (versus the conventional or personal domains) as absolute (e.g., Nucci & Turiel, 1978; Smetana, 1981; Turiel, 1983; see Killen & Smetana, 2006, for review). In addition, Nobes, Panagiotaki, and Pawson (2009; see also Schleifer, Shultz, & Lefebvre-Pinard, 1983) have recently suggested that it may be difficulty with the concept of negligence, rather than intention, that underlies young children’s classic failures: they may assume (even if told otherwise) that bad outcomes result from negligence or carelessness and
are blameworthy as a result (see also Walster, 1966; Yuill, 1984). Critically, in a recent study that explicitly distinguished intention and care, all children (3 years was the youngest tested) treated intention as most important in assigning praise or blame (Nobes et al., 2009). Together, these findings are difficult to interpret within Piaget’s developmental progression of objective-to-subjective morality: young children do not seem to simply believe that wrong = breaking X rule or causing Y outcome. How can these divergent results, and the conclusions about moral development that follow, be reconciled?

Originally claimed by Piaget and Inhelder (1962), a large amount of cognitive development research suggests that morality is but one of several domains in which young children are “objective” in their judgments. Within the broader social domain, for instance, “Theory of Mind” (ToM) research suggests that while very young children grasp that actions are driven by mental states, they may confuse their own mental states with others’ (e.g., Repacholi & Gopnik, 1997), and have difficulty understanding the unique roles played by distinct mental states, such as intention versus desire (e.g., Astington, 1991, 1993; Feinfield, Lee, Flavell, Green, & Flavell, 1999; Moses, 1993). In addition, it has been repeatedly demonstrated that children under 4.5 years have difficulty with representational mental states such as ignorance and false belief, tending to focus on what they know about the objective state of the world (see Wellman, Cross, & Watson, 2001, for review; see Astington & Gopnik, 1991; Gopnik & Seiver, 2009, for general reviews of the development of Theory of Mind). Indeed, individual children’s performance on Theory of Mind tasks is related to whether they privilege mental states in their moral judgments, suggesting the processes are intimately linked (e.g., Chandler et al., 2001; Killen et al., 2011).

Evidence for developmental change in mentalizing abilities notwithstanding, there is now a large body of work suggesting that preverbal infants interpret agents’ object-directed actions in terms of their goal-relevant properties, suggesting that at least some forms of mentalizing are present from early in development. By 3–5 months of age infants pay more attention to events demonstrating mentalistic changes (changing one’s mind) over physical changes (changing one’s motion; see Woodward, 2009, for review). In both attentional and behavioral paradigms, infants have been shown to privilege mental states in cases in which goals are not met or conflict with outcome; for instance, in failed attempts and accidents (e.g., Behne, Carpenter, Call, & Tomasello, 2005; Brandone & Wellman, 2009; Carpenter, Akhtar, & Tomasello, 1998; Hamlin, Hallinan, & Woodward, 2008; Hamlin, Newman, & Wynn, 2009; Legerstee & Markova, 2007; Marsh, Stavropoulos, Nienhuis, & Legerstee, 2010). In addition, it has recently been documented that infants can incorporate others’ false beliefs into their action interpretations (for review, see Baillargeon, Scott, & He, 2010; see also Buttelman, Carpenter, & Tomasello, 2009; Kovacs, Teglas, & Endress, 2010; Luo, 2011; Southgate, Senju, & Csibra, 2007); false belief understanding has long been considered the critical test case for a mature Theory of Mind.

How can results suggesting that infants mentalize, privilege intention over outcome in a variety of ways, and are sensitive to false beliefs, be reconciled with classic demonstrations of Theory of Mind failures in young children both within and outside of the moral domain? Critics of infant work claim that the use of attentional paradigms is problematic, noting that infants’ attention may be drawn to one event over another for a host of reasons that might not reflect true competence (e.g., Cohen & Cashon, 2006; Haith & Benson, 1998; Kagan, 2008; Perner & Ruffman, 2005; Rakison, 2007; Ruffman & Perner, 2005; Ruffman, Taumoepeau, & Perkins, 2012). And yet, a significant minority of infant Theory of Mind research has utilized non-attentional methodologies such as imitation and emotional reactivity, and has demonstrated the same competencies. One common explanation for the dichotomy between studies using infants and children is that typical mentalizing tasks for children are too cognitively-demanding for the youngest children tested; that is, that young children’s failures may be due to immature domain-general abilities rather than domain-specific ones. For example, the common “Sally-Anne” false belief task (e.g., Wimmer & Perner, 1983) requires children to keep track of multiple individuals’ perspectives (even when some individuals leave the scene), to inhibit a pre-potent response or heuristic to respond based on reality, to engage in counterfactual reasoning, and to respond to directly-elicited verbal questioning. To demonstrate that typical “elicited-response” task demands are to blame for young children’s failures, researchers have developed various “spontaneous-response” tasks that reduce demands on children’s domain-general cognitive capacities. Indeed, infants and younger children have been shown to pass mentalizing tests tapping spontaneous responses such as attentional methodologies (again, the value of such tasks is debated), imitation, play-acting, emotional reactions, and responses to implied – rather than explicit – verbal cueing (e.g., Baillargeon et al., 2010; Bloom & German, 2000; Leslie & Polizzi, 2002; Rubio-Fernandez & Geurts, 2013; Scott, He, Baillargeon, & Cummins, 2012).

Although there is now increasing evidence from spontaneous-response that very young children can interpret agentic action mentalistically, very young children may not interpret all actions in this way. For one thing, morally relevant actions are usually valenced: moral actors have positive or negative intentions, and those actions tend to result in positive or negative outcomes for others; perhaps valence distracts very young children and disrupts their mentalistic processing. Evidence against this possibility comes from several studies using an “unwilling versus unable” paradigm, in which experimenters are either unwilling or unable to give infants something they desire (e.g., Behne et al., 2005; Dunfield & Kuhlmeier, 2010; Legerstee & Markova, 2007; Marsh et al., 2010; see also Call, Hare, Carpenter, & Tomasello, 2004). That is, unwilling experimenters have a mean intention (to deny infants something they desire) and are effective in bringing out a negative outcome (infants do not get toy/attention); whereas unable experimenters have a nice intention (to satisfy infants’ desire) but are ineffective in bringing out positive outcome (infants do not get toy/attention). Critically, although the outcome for participants is always negative, infants become more agitated while interacting with an unwilling...
than an unable experimenter, suggesting they distinguished their actions based solely on mental states.

In a recent adaptation of the unwilling/unable methodology, Dunfield and Kuhlmeier (2010) found that 21-month-old toddlers selectively helped unable over unwilling experimenters, presumably because they liked the unable experimenter and/or disliked the unwilling one (though perhaps it was because they noticed the unable experimenter was ineffective and decided she could use some help). In contrast, toddlers helped able (nice intention, effective at bringing out positive outcome) and unable (nice intention, ineffective at bringing about positive outcome) experimenters at equal rates. That is, despite the potentially strong pull of personal interest toward individuals who effectively provide personal benefits, and although unable experimenters may generally be viewed as needing help, outcome-valence did not influence toddlers’ helping behaviors in this study.

Such results suggest that infants can interpret valenced prosocial and antisocial behaviors mentalistically. Yet, it is presently unclear whether there is any relationship between infants’ responses in the unwilling/unable paradigm and older children’s moral judgments. The term “moral” tends to refer only to situations in which evaluators are not personally involved, in which “interest or revenge or envy perverts not our disposition” (Hume, 1898); in studies using the unwilling/unable paradigm infants and toddlers were always the direct recipients of prosocial and antisocial acts. To get at this question, another methodological paradigm has examined whether infants evaluate others for their third-party prosocial and antisocial behaviors; that is, actions in which infants are mere bystanders and have no immediate personal interest. This has been dubbed the “Helper/Hinderer” paradigm.

The first studies using the Helper/Hinderer paradigm to examine infants’ evaluations of third-party prosocial and antisocial acts involved a character (a “Protagonist”) who demonstrated an unfulfilled goal to reach the top of a steep hill (Hamlin et al., 2007; adapted from Kuhlmeier, Wynn, and Bloom, 2003); see also Premack & Premack, 1997). On alternating events, a “Helper” facilitated the Protagonist’s goal by pushing him up the hill, and a “Hinderer” blocked the Protagonist’s goal by pushing him down the hill. Six- and 10-month-old infants were habituated to these helping and hindering events, and their evaluations of the Helper and Hinderer were subsequently examined via a preferential-reaching method. Both age groups robustly preferred the Helper to the Hinderer (Hamlin et al., 2007; but see critique by Scarf, Imuta, Colombo, and Hayne, 2012) and response by Hamlin, Wynn, and Bloom (2012)). Subsequent studies have conceptually replicated this effect with additional Helper/Hinderer scenarios, involving goals such as retrieving a dropped object, opening a box to get a toy, and lifting an obstacle blocking a preferred-object (e.g., Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2010; Hamlin et al., 2007). Across studies and from 3 months of age, infants robustly preferred (looked more to or chose) Helpers over Hinderers, suggesting the tendency to evaluate others for their third-party prosocial and antisocial acts emerges within the first year after birth. Critically, for several of these paradigms new groups of infants were shown puppets enacting the *exact same behaviors* on an inanimate object (Hamlin & Wynn, 2011), or on a social character whose physical behaviors were identical but whose goal was ambiguous (Hamlin Ullman Tenenbaum et al., 2013). Infants in these control conditions did not prefer either puppet, suggesting that infants’ evaluations of Helpers and Hinderers are based on their social, goal-relevant actions rather than on performing particular physical actions or bringing about particular end-states.

Looking more at one character than another, or choosing one character over another, is presumably not a perfect measure of “evaluation”; however, one’s toolkit for determining young infants’ social attitudes is limited to behaviors they can perform. Importantly, other studies have demonstrated that younger infants’ visual preferences are consistent with their manual ones (Hamlin et al., 2010), and that younger infants’ manual preferences are consistent with older infants’ active social behaviors: 21-month-olds give treats to Helpers but take them from Hinderers (e.g., Hamlin, Wynn, Bloom, & Mahajan, 2011), and 16-month-olds match the food preferences of Helpers but not of Hinderers (e.g., Hamlin & Wynn, 2012). In addition, infants in the second year after birth prefer those who distribute objects fairly versus unfairly (e.g., Geraci & Surian, 2011; see also Sloane, Baillargeon, & Premack, 2012), suggesting that at least by the second year infants’ social evaluations are not limited to notions of helping and hindering per se, but apply to a wide swathe of morally-relevant behaviors.

Of course, infants’ tendency to socially evaluate others, even for those behaviors adults might consider “moral” (helping) or “immoral” (hindering), is not necessarily part of an emerging system of moral evaluation. Indeed, it is as of yet unclear even whether infants’ evaluations reflect the same criteria as adults’ do. Before the relationship between infant social evaluation and mature moral evaluation can be considered, then, it seems critical to determine exactly what it is that infants “like” about Helpers and/or “dislike” about Hinderers. One possibility is that infants positively evaluate anyone who facilitates a goal and negatively evaluate anyone who prevents a goal. That is, adults tend not to praise individuals who, despite facilitating some individual’s goal, provide getaway cars in bank robberies or who let their children eat candy for every meal. But infants may not see the distinction between “helpful” and “good” or between “unhelpful” and “bad”; if not their evaluations would be quite unlike adults’.

Recent research suggests one way in which infants do distinguish helpful from good by the end of the first year: rather than always preferring goal-facilitators and avoiding goal-blockers, 8-month-olds prefer those who hinder the goals of characters who have previously hindered others (e.g., Hamlin et al., 2011), and 9-month-olds prefer those who hinder the goals of characters who are dissimilar to themselves (e.g., Hamlin, Mahajan, Liberman, & Wynn, 2013). But nuanced preferences like these do not necessarily speak to whether infants engage in *mentalistic* social evaluation: though preferring someone who hinders a disliked other could be based on what the Hinderer intended to accomplish by hindering (for instance, providing deserved punishment), it need not be. Rather, infants...
might assess which individuals are their “friends,” and friends include the enemies of their enemies (see Heider, 1958). Together, heuristics like these could explain previous findings via analyses of complex, but nonetheless objective and non-mentalistic, aspects of prosocial and antisocial acts.

2. The current studies

The current studies explored whether infants incorporate mental states into their third party social evaluations by examining their preferences for puppets who try but fail to help or hinder another’s goal. Infants viewed puppet events in which a Protagonist (cow wearing yellow shirt) demonstrated an unfulfilled goal to open a clear box with a toy inside (as in Hamlin and Wynn (2011)). Two additional puppets (pigs, one wearing a green shirt, one wearing a blue shirt) then demonstrated either a positive intention to help the Protagonist get his goal by joining in the struggle to open the box, or a negative intention to prevent the Protagonist from getting his goal by slamming it shut. Across several conditions, some Pigs successfully helped or hindered (“Successful Pigs”), whereas other Pigs failed to help or hinder (“Failed Pigs”). Thus, each Pig + Protagonist interaction was defined by (1) the valence of the Pig’s intention toward the Protagonist (positive to help the Protagonist open the box, negative to prevent the Protagonist from opening the box), and (2) whether the Protagonist achieved his goal in the end (a positive outcome for the Protagonist) or not (a negative outcome for the Protagonist).¹ From these two dimensions, 4 character-types emerge: (1) “Successful Helpers” who try to help the Protagonist achieve his goal and do indeed help; (2) “Successful Hinderers” who try to prevent the Protagonist from achieving his goal and do indeed hinder; (3) “Failed Helpers” who try to help the Protagonist but do not help; and (4) “Failed Hinderers” who try to prevent the Protagonist from achieving his goal but do not hinder. During Successful events the Pig’s intention and the outcome for the Protagonist are of the same valence; during Failed events the Pig’s intention and the outcome for the Protagonist are of opposite valence.

Experiments 1–3 examined infants’ relative preferences for Successful and Failed Pigs in various combinations of 2 (see Table 1 for a summary of conditions). Experiment 1 examined whether infants can distinguish Pigs based solely on their intention to help or hinder the Protagonist, even when the Protagonist experiences the same outcome at the end of each interaction. In 1A, a Successful Helper (Pig intention to help = positive, Protagonist achieves goal = positive) was pitted against a Failed Hinderer (Pig intention to hinder = negative, Protagonist achieves goal = positive). In 1B, a Failed Helper (Pig intention to help = positive, Protagonist fails to achieve goal = negative) was pitted against a Successful Hinderer (Pig intention to hinder = negative, Protagonist fails to achieve goal = negative). Experiment 2 examined whether infants privilege Pig intention or Protagonist outcome when they both differ and are in conflict: a Failed Helper (Pig intention to help = positive, Protagonist fails to achieve goal = negative) was pitted against a Failed Hinderer (Pig intention to hinder = negative, Protagonist achieves goal = positive). Finally, Experiment 3 examined whether infants distinguish Pigs when the Protagonist experiences different outcomes at the end of each interaction but Pigs’ intentions are of the same valence: In 3A, a Successful Helper (Pig intention to help = positive, Protagonist achieves goal = positive) was pitted against a Failed Helper (Pig intention to help = positive, Protagonist fails to achieve goal = negative). In 3B, a Failed Hinderer (Pig intention to hinder = negative, Protagonist achieves goal = positive) was pitted against a Successful Hinderer (Pig intention to hinder = negative, Protagonist fails to achieve goal = negative). To summarize, in Experiment 1 Pigs’ intentions were of different valence while the Protagonist’s outcomes were of the same valence, in Experiment 2 both Pig’s intentions and the Protagonist’s outcomes were of different valence (opposed each other), and in Experiment 3 Pigs’ intentions were of the same valence while the Protagonist’s outcomes were of different valence.

In all Experiments, infants’ evaluations were examined by having them choose between the 2 Pigs they had seen act on the Protagonist during habituation (e.g., Geraci & Surian, 2011; Hamlin & Wynn, 2011; Hamlin et al., 2007).

Table 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Protagonist outcome</th>
<th>Pig intentions</th>
<th>Protagonist outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A: Successful Helper versus Failed Hinderer</td>
<td>Different; +</td>
<td>Same; +</td>
<td></td>
</tr>
<tr>
<td>1B: Failed Helper versus Successful Hinderer</td>
<td>Different; +</td>
<td>Same; −</td>
<td>−</td>
</tr>
<tr>
<td>2: Failed Helper versus Failed Hinderer</td>
<td>Different; +</td>
<td>Different; −</td>
<td>−</td>
</tr>
<tr>
<td>3A: Successful Helper versus Failed Hinderer</td>
<td>Same; +</td>
<td>Different; −</td>
<td>−</td>
</tr>
<tr>
<td>3B: Successful Hinderer versus Failed Hinderer</td>
<td>Same; −</td>
<td>Different; +</td>
<td>−</td>
</tr>
<tr>
<td>4: Coordinated versus Slammer Pig</td>
<td>Same; Both neutral</td>
<td>Same; −</td>
<td>−</td>
</tr>
<tr>
<td>5: Successful Helper versus Successful Hinderer</td>
<td>Different; +</td>
<td>Different; −</td>
<td>−</td>
</tr>
</tbody>
</table>

¹ One could also consider outcome-valence from the perspective of the Helpers/Hinderers (that is, positive if their goal to help or to hinder is achieved; negative if their goal to help or to hinder is not achieved). Throughout this paper “outcome-valence” will always refer to the valence from the perspective of the Protagonist: getting the box open fulfills the Protagonist’s goal and is positive; not getting the box open fails to fulfill the Protagonist’s goal and is negative. This is most consistent with how intentions and outcomes are typically conceived of in the social and moral evaluation world; for instance, Bobby might intend to hug Polly (a positive intention) but end up hitting her in the face (a negative outcome for Polly).
2010, 2011). Both 5- and 8-month-olds were tested in each Experiment; these ages were chosen due to previous research showing (1) that infants prefer Successful Helpers to Successful Hinderers in this same box-opening situation by 5 months of age (Hamlin & Wynn, 2011) and (2) that 8-month-olds show sensitivity to context in their social evaluations (preferring Hinderers of Hinderers) but 5-month-olds do not (preferring Helpers of Hinderers; e.g., Hamlin et al., 2011).

In addition, to determine whether adults interpreted the puppet events as designed, adult participants (n = 10 in each condition of Experiments 1, 2, and 3) were shown videos of the puppet events and asked to write a short sentence depicting what they saw. In addition, after adults viewed both events they were asked to identify (1) whom they preferred, and (2) who was nicer.

3. Experiment 1A and 1B: do infants evaluate others based on mental states when outcomes are uninformative?

3.1. Method

3.1.1. Participants

3.1.1.1. Infant participants. All infants were healthy and full-term, recruited from a midsize North American city. In Experiment 1A, sixteen 5-month-olds (8 boys; mean age = 4;28), and sixteen 8-month-olds (8 boys; mean = 7;24) chose between a Successful Helper and a Failed Hinderer; in Experiment 1B sixteen 5-month-olds (8 boys; mean = 5;0) and sixteen 8-month-olds (8 boys; mean = 7;28) chose between a Failed Helper and a Successful Hinderer. Five additional 5-month-olds and 10 additional 8-month-olds participated but were excluded from analyses due to procedure error (2 five-month-olds/3 eight-month-olds), fussiness (2/4), failure to choose either puppet (1/1) and family interference (0/2).

3.1.1.2. Adult participants. Twenty adults (10 male, 10 female) were recruited from non-psychology campus buildings (communal study spaces) to participate in a 5-min study in exchange for candy. Subjects included mostly undergraduates, some graduate students, and some adult building staff. Ten (5 male, 5 female) were shown videos of a Successful Helper and a Failed Hinderer; 10 (5 male, 5 female) were shown videos of a Failed Helper and a Successful Hinderer. One additional adult participated in condition 1B but was not included in the final sample as he was part of a group of 5 individuals who refused to be tested individually and discussed the videos amongst themselves.

3.1.2. Procedures – infants

Infants viewed puppet events from their parent’s lap at the end of a long table surrounded on 3 sides by black curtains; a fourth curtain could be raised and lowered to occlude the puppet stage between events. A puppeteer experimenter performed puppet events by placing her hands underneath the rear curtain; she wore a black shirt to cover her arms and was entirely invisible to infants. Parents were instructed to sit quietly with their infants and to not attempt to influence them in any way. Infants’ attention was recorded by an independent coder peeking through a hole in the curtain on the infants’ right side; she could not see the puppet events.

3.1.2.1. Habituation. Events are depicted in Fig. 1, and at http://cic.psych.ubc.ca/Failed_Attempts_Videos.html. Regardless of condition or event type, all events began identically: a curtain occluder rose to reveal a clear box containing an elephant toy; two pigs wearing a green and blue shirt sat at the rear corners of the stage.

At the start of each trial, the Protagonist (cow with yellow shirt) 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, apparently ‘looking’ at the toy inside the box, then jumped onto the front corner of the box lid. He then attempted to open the box. On each of his first and second attempts, the Protagonist lifted and shook the lid to indicate a struggle, dropped it shut, and sat up to face the infants. During attempts 3 and 4, he lifted and shook the lid and lowered it back down, all while continuously holding it (procedure identical to Hamlin and Wynn (2011)). On the Protagonist’s 5th attempt, the Pig resting at the opposite rear-corner of the stage intervened. The following describes what happened following this point for each event type:

During Successful Helper events (Pig intention to help = positive, Protagonist achieves goal = positive; Fig. 1, line 1, Video S1), the Successful Helper moves forward and grasps the box lid, and together opens the box with the Protagonist. The Protagonist lies down inside the box on top of the toy (as though acquiring it) and the Successful Helper runs off-stage.

During Failed Helper events (Pig intention to help = positive, Protagonist fails to achieve goal = negative; Fig. 1, line 2, Video S2), the Failed Helper moves forward and grasps the box lid (as in Successful Helper events), and together with the Protagonist makes three more failed attempts to open the box: both puppets lift, shake, and lower the box lid three times while holding it continuously. At the end of the third attempt, both characters “give up” and release the lid. The Protagonist lies down next to the box without the toy, and the Failed Helper runs off-stage.

During Successful Hinderer events (Pig intention to hinder = negative, Protagonist achieves goal = negative; Fig. 1, line 3, Video S3), the Successful Hinderer moves forward and jumps onto the opposite corner of the lid, slamming it shut. The Protagonist lies down next to the box without the toy, and the Successful Hinderer jumps off of the box and runs off-stage.

During Failed Hinderer events (Pig intention to hinder = negative, Protagonist achieves goal = positive; Fig. 1, line 4, Video S4), the Failed Hinderer moves forward and jumps onto the lid, slamming it shut. However, after the Protagonist lies down next to the box as in Successful Hinderer events, the Failed Hinderer jumps off the box and pauses next to it. The Protagonist then sits up and smoothly opens the lid on his own, grasping the toy as in Successful Helper events. The Failed Hinderer runs off-stage.
Infants in Experiment 1A viewed Successful Helper events and Failed Hinderer events in alternation; that is, the Pigs differ in intention but the Protagonist always achieves its goal. Infants in Experiment 1B viewed Failed Helper and Successful Hinderer events in alternation; that is, the Pigs differ in intention but the Protagonist never achieves its goal. During all events, all action paused once the intervening puppet left the stage. Infants' attention was recorded from this point by the online coder using the custom coding program JHab (Casstevens, 2007), until (1) infants looked away for 2 consecutive seconds, or (2) 30 s elapsed. Infants were habituated to alternating events until either (1) their looking time on three consecutive trials was less than half of their looking time on the first three trials, or (2) they had seen 14 total trials.

### Choice
Following habituation, parents rotated their chairs clockwise 90° from the stage, positioned their infants sitting straight ahead at the front of their laps, and closed their eyes. The online coder from habituation (who had not seen the puppet events and was blind to the puppets' identities) held the puppets up in front of the infants, approximately 30 cm apart and initially out of reach. Infants were required to look at both puppets and back to the experimenter, once they did so the puppets were moved within reach. Choice was determined online by this experimenter as the first puppet infants contacted with a visually-guided reach (that is, a touch to a puppet that was preceded by a look to that puppet). Patterns of choice within condition were compared to chance (.5) via binomial tests, and across condition via Fisher's Exact tests. All reported p-values are two-tailed.

### Procedures – adults
Once an adult agreed to participate, an experimenter told him or her “You are going to watch a puppet show with a cow puppet and two pig puppets in green and blue shirts. Afterward I will ask you some questions about what you saw.” Participants were shown two videos in succession; each was a recording of the very same puppet events shown to infants in each condition. The experimenter controlled video play from behind the laptop via remote control, allowing her to remain blind to condition. She determined condition assignment during the procedure via random-number generator. If any participant requested to see the videos again, they were played a second time only if the request came before the participant had been asked to identify a preferred puppet. If the replay request came after this point, the request was denied and the procedure continued.

After adult participants watched each event of their condition, they were asked to “please write one sentence depicting what you saw in the show.” After they had written a sentence on the paper provided, they were asked to

---

**Fig. 1.** Stimuli presented to infants/adults in Experiments 1–3, 5. Line 1: Successful Helper Event in Experiments 1A, 3A, and 5. (a) Protagonist tries to open box; (b) Successful Helper moves forward and grasps box lid; (c) protagonist and Successful Helper open box lid together; (d) protagonist lays down on toy inside box, Successful Helper runs offstage. Line 2: Failed Helper Event in Experiments 1B, 2, and 3A. (a) Protagonist tries to open box; (b) Failed Helper moves forward and grasps box lid; (c) protagonist and Failed Helper struggle with lid; (d) protagonist lays down next to box, Failed Helper runs offstage. Line 3: Successful Hinderer Event in Experiments 1B, 3B, and 5. (a) Protagonist tries to open box; (b) Successful Hinderer moves forward and jumps on top of box; (c) protagonist lays down next to box; (d) Successful Hinderer runs offstage. Line 4: Failed Hinderer Event, Experiments 1A, 2, and 3B. (a) Protagonist tries to open box; (b) Failed Hinderer moves forward and jumps on top of box; (c) Failed Hinderer jumps off box and Protagonist opens box alone; (d) protagonist lays down on toy in box, Failed Hinderer runs offstage.
indicate “of the two, which puppet did you like the most?” by writing down a Pig’s t-shirt color. If, following the preference question, participants asked whether they could choose neither or both puppets as their preferred puppet, they were asked to “please pick one”. Following the preference measure, participants were asked “and which puppet did you think was nicer? You can pick whoever you’d like, both, or neither.” The experimenter subsequently transcribed participants’ written descriptions and choice responses, remaining blind to condition during transcription.

The “nicest” measure was included in addition to the “like most” measure because a few pilot subjects asked whether they could identify “neither” puppet as their choice following the “like most” question. To keep the adult procedure maximally similar to the infant procedure, adults were forced to choose a puppet for the “liking” question. However, it was conceivable that when forced to choose a preferred puppet, adults would identify a dimension on which to distinguish the puppets whether or not they found that dimension most meaningful or relevant. Thus, the niceness question was included to help inform whether or not participants considered the puppet they liked to be nice, based on the possibility that this might be true in some conditions and not in others. A second motivation for asking participants which puppet was nicer was that a few pilot participants laughed upon seeing the hindering events. This was not a terribly surprising reaction: Events were designed for preverbal infants and involved cute puppets, and the hindering actions were only mildly negative by way of pigs behaving antisocially. Of course, finding the Hinderer’s behavior humorous, despite recognizing that the Hinderer was a few pilot participants laughed upon seeing the hindering events. This was not a terribly surprising reaction: Events were designed for preverbal infants and involved cute puppets, and the hindering actions were only mildly negative by way of pigs behaving antisocially. Of course, finding the Hinderer’s behavior humorous, despite recognizing that the Hinderer was mildly negative, this concern is far less of an issue with adult participants.

3.2. Results

3.2.1. Habituation

A summary of infants’ attention for every experiment is available in Table 2. Including non-habituators (who saw 14 events; these infants’ choices never differed from habituators and were included here and in all reported data), both 8- and 5-month-olds saw an average of 9 habituation trials in the Successful Helper/Failed Hinderer condition (1A). In the Failed Helper/Successful Hinderer condition (1B), 8-month-olds were shown an average of 9 and 5-month-olds were shown an average of 10 habituation trials; this did not differ by age (t(30) = −1.46, p = .15; η² = .07). Two 8-month-olds and three 5-month-olds in the Successful Helper/Failed Hinderer condition and two 5-month-olds in the Failed Helper/Successful Hinderer condition failed to habituate in 14 trials.

Across both conditions, ANOVAs revealed that total attention to the first 6 habituation events differed by age, with 5-month-olds looking less than 8-month-olds (F(1,62) = 6.17; p < .05; η² = .09). Total attention over the first 6 habituation events (collapsed across event type) did not differ by condition within or across ages (p’s > .12), and attention to individual event types did not differ within or across age or condition (p’s > .33).

3.2.2. Choice – infants

Infants’ choices are depicted in Fig. 2. Eight-month-olds in Experiment 1 preferred the puppet with the positive intention, regardless of the valence of the outcome the Protagonist experienced. Specifically, in 1A they preferred the Successful Helper to the Failed Hinderer (14 of 16 infants, binomial p < .005), and in 1B they preferred the Failed Helper to the Successful Hinderer (14 of 16 infants, binomial p < .005). Collapsed across conditions, 28 of 32 infants chose the puppet with the positive intention to help (binomial p < .00005); there was no effect of Protagonist-outcome (Fisher’s Exact Test, p = 1), or Pig-success/failure (binomial p = 1) on infants’ choices.

Unlike 8-month-olds, 5-month-olds failed to distinguish the characters in either condition; they chose equally between a Successful Helper and a Failed Hinderer (10/16 chose Helper, binomial p = .45) and between a Failed Helper and a Successful Hinderer (8/16 chose the Helper, binomial p = 1). Across conditions, there were no effects of Pigs’-intention valence (binomial p = .60), Protagonist-outcome valence (Fisher’s Exact Test, p = .71), or Pig-success/failure (binomial p = .60) on choice.

Collapsed across condition, infants’ choice patterns differed significantly by age (Fisher’s Exact Test, p < .05), reflecting that 8-month-olds distinguished the characters and 5-month-olds did not. Univariate ANOVAs on infants’ choices (coded as 1 if infant chose the positively-intended Pig, 0 if chose the negatively-intended Pig) with attention over the first 6 habituation events as a covariate revealed no effects of attention on choice within or across age groups or conditions (all Fs < 1.3; all p’s > .25).

3.2.3. Choice – adults

Adults’ written descriptions are listed in the Appendix; choices are in Fig. 3. In Experiment 1A, 9 adults “liked” the

---

Footnote:

2 Infant participants were required to make a choice to remain in the sample. Two infants in Experiment 1 failed to choose either puppet, but failure to choose in infants might indicate mere shyness as opposed to equal puppet liking; this concern is far less of an issue with adult participants.
Table 2
Infants’ average total looking times during habituation, etc. “Better” refers to the event hypothesized to be more-positive (relative to condition: could refer to intention, outcome, physical action, etc.); “worse” refers to the less-positive event (relative to condition). All numbers in parentheses are standard errors.

<table>
<thead>
<tr>
<th></th>
<th>Age group</th>
<th># Habituation trials</th>
<th>Attention to first 3 habituation trials</th>
<th>Attention to last 3 habituation trials</th>
<th># Infants who did not habituate in 14 trials</th>
<th>Attention to first 3 Successful Helper trials</th>
<th>Attention to first 3 Successful Hinderer trials</th>
<th>Attention to first 3 Failed Helper trials</th>
<th>Attention to first 3 Failed Hinderer trials</th>
<th>Attention to first 3 Coordinated trials</th>
<th>Attention to first 3 Slammer trials</th>
<th>Attention to “better” trials</th>
<th>Attention to “worse” trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment A</strong></td>
<td>8</td>
<td>8.5 (.69)</td>
<td>26.03 (3.47)</td>
<td>12.57 (1.38)</td>
<td>2</td>
<td>20.43 (2.39)</td>
<td>24.24 (3.60)</td>
<td></td>
<td></td>
<td>20.43 (2.39)</td>
<td>24.24 (3.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.31 (.81)</td>
<td>15.75 (2.97)</td>
<td>7.93 (.88)</td>
<td>3</td>
<td>12.33 (1.70)</td>
<td>13.76 (2.69)</td>
<td></td>
<td></td>
<td>12.33 (1.70)</td>
<td>13.76 (2.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 1B</strong></td>
<td>8</td>
<td>8.69 (.69)</td>
<td>30.15 (4.35)</td>
<td>10.96 (1.57)</td>
<td>0</td>
<td>25.47 (4.0)</td>
<td>25.28 (3.85)</td>
<td></td>
<td></td>
<td>25.28 (3.85)</td>
<td>25.47 (4.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10.25 (.81)</td>
<td>21.37 (4.34)</td>
<td>8.91 (1.75)</td>
<td>2</td>
<td>20.83 (4.92)</td>
<td>18.35 (2.99)</td>
<td></td>
<td></td>
<td>18.35 (2.99)</td>
<td>20.83 (4.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td>8</td>
<td>7.75 (.56)</td>
<td>25.62 (2.95)</td>
<td>10.01 (1.37)</td>
<td>1</td>
<td>19.51 (2.36)</td>
<td>21.85 (2.32)</td>
<td></td>
<td></td>
<td>19.51 (2.36)</td>
<td>21.85 (2.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>8.63 (.67)</td>
<td>22.51 (3.26)</td>
<td>9.29 (1.09)</td>
<td>0</td>
<td>17.93 (2.95)</td>
<td>21.91 (3.89)</td>
<td></td>
<td></td>
<td>17.93 (2.95)</td>
<td>21.91 (3.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 3A</strong></td>
<td>8</td>
<td>8.25 (.57)</td>
<td>32.73 (4.03)</td>
<td>12.42 (2.03)</td>
<td>0</td>
<td>26.15 (3.34)</td>
<td>28.03 (3.63)</td>
<td></td>
<td></td>
<td>26.15 (3.34)</td>
<td>28.03 (3.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10.63 (.73)</td>
<td>20.55 (3.97)</td>
<td>11.70 (3.21)</td>
<td>3</td>
<td>21.28 (3.63)</td>
<td>19.98 (4.41)</td>
<td></td>
<td></td>
<td>21.28 (3.63)</td>
<td>19.98 (4.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 3B</strong></td>
<td>8</td>
<td>9.63 (.74)</td>
<td>22.90 (2.83)</td>
<td>10.32 (1.76)</td>
<td>2</td>
<td>20.86 (2.59)</td>
<td>21.72 (2.58)</td>
<td></td>
<td></td>
<td>21.72 (2.58)</td>
<td>20.86 (2.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>8.88 (.83)</td>
<td>33.54 (5.54)</td>
<td>15.78 (3.18)</td>
<td>3</td>
<td>26.33 (4.21)</td>
<td>28.87 (4.50)</td>
<td></td>
<td></td>
<td>28.87 (4.50)</td>
<td>26.33 (4.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 4</strong></td>
<td>8</td>
<td>8.56 (.60)</td>
<td>26.52 (2.84)</td>
<td>10.40 (1.40)</td>
<td>1</td>
<td>25.11 (4.28)</td>
<td>24.36 (3.73)</td>
<td></td>
<td></td>
<td>25.11 (4.28)</td>
<td>24.36 (3.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 5</strong></td>
<td>5</td>
<td>10.38 (.64)</td>
<td>20.85 (2.98)</td>
<td>9.71 (1.29)</td>
<td>2</td>
<td>21.49 (3.05)</td>
<td>20.07 (3.31)</td>
<td></td>
<td></td>
<td>21.49 (3.05)</td>
<td>20.07 (3.31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Helper and 1 “liked” the Failed Hinderer (binomial $p < .05$), replicating 8-month-olds’ pattern of reaching. Nine of 10 identified the Successful Helper as “nicer,” and the remaining participant indicated that neither puppet was nicer; no adults identified the Successful Hinderer as nicer (binomial $p < .001$). Although written descriptions suggest that 10/10 participants appropriately analyzed the behaviors (for instance, 9/10 adults specifically used the word “help” in their description), the participant who indicated that neither puppet was nicest seemed to view the Failed Hinderer’s action as facilitating the box opening as well: “The pigs help the cow open the box. The first one pops it open, and the second helps the cow pry it open” (italics added).

In Experiment 1B, 8 adults liked the Failed Helper, and 2 liked the Hinderer (binomial $p = .10$), marginally replicating 8-month-olds’ reaches. The two subjects who stated they liked the Hinderer nonetheless described the interactions appropriately; one said “The first pig can’t help, and the other jumps on the box so the cow can’t get inside,” and the other said “A pig tries to help the cow open the box, and another pig jumps on the box.” All 10 adults identified the Failed Helper as nicer (binomial $p < .00005$), and 9 of 10 subjects identified the Failed Helper as trying or being unable to help in their written responses.

3.3. Discussion

Results from Experiment 1 suggest that infants can evaluate others for their third party actions based solely on intention by 8 months of age, although not yet by 5 months. Adults’ choices and written responses suggest that they appropriately interpreted the puppet events, and that their preferences were consistent with infants’. These results support the claim that infants’ social evaluations are sensitive to intention; however, there are at least two reasons to be cautious of a mentalistic interpretation of infants’ reaching patterns in Experiment 1. First, because outcomes were equated in each condition in Experiment 1 it is currently unknown whether infants’ evaluations merely utilize mental states when outcomes are equated (as do 3-year-olds; e.g., Leon, 1980), or if they also privilege mental states when outcomes differ and oppose intention. Second, it is also possible that infants in Experiment 1 simply ignored the Failed Pigs in each condition, and based their evaluations solely on the behaviors of Successful Pigs. Several previous studies have shown that infants selectively approach Successful Helpers and avoid Successful Hinderers (e.g., Hamlin et al., 2007, 2011); these tendencies could explain 8-month-olds’ choices in Experiment 1 with no reference to mental states at all.

Unlike 8-month-olds, 5-month-olds in Experiment 1 failed to distinguish the characters in either condition. This failure may have resulted from several cognitively distinct difficulties. First, 5-month-olds may have trouble mentalistically interpreting failed attempts of any kind, and therefore cannot distinguish puppets based on distinct intentions. Indeed, although some previous work has shown that 3- to 8-month-olds appropriately react to certain cases of failed attempts (e.g., Hamlin et al., 2008, 2009; Legerstee & Markova, 2007; Marsh et al., 2010), other studies have documented failures in infants under 9–10 months of age (e.g., Behne et al., 2005; Brandone & Wellman, 2009), suggesting that the ability may be relatively unstable. In addition, there is some evidence that infants under 12 months fail to interpret even successful box-opening behaviors mentalistically (in terms of the ultimate goal of the object in the box, e.g., Woodward & Sommerville, 2000). Together, these results suggest that understanding a failed attempt to intervene on someone else’s failed attempt to open a box in terms of mental states may be quite difficult for young infants. An additional possibility underlying 5-month-olds’ failures is that 5-month-olds may simply have difficulty with conflicting information of any kind (e.g., Diamond, 1990; see Zelazo, Muller, Frye, & Marcovitch, 2003, for review). Information conflict might come from many sources, but one potential source is a situation in which mental states oppose outcomes, as is the case in the current studies. Indeed, in a previous study 5-month-olds failed to incorporate information that conflicted across multiple events into their evaluations (Hamlin et al., 2011); as the current studies require interpreting conflicting information within the same event, they might be especially taxing for young infants.

To further examine the various alternatives for infants’ choices in Experiment 1, 8- and 5-month-olds in Experiment 2 viewed Pigs whose intention valence was pitted against that of the outcome the Protagonist experienced: a Failed Helper (Pig intention to help = positive, Protagonist fails to achieve goal = negative) versus a Failed Hinderer (Pig intention to hinder = negative, Protagonist achieves goal = positive). Critically, this contrast helps to

![Fig. 2. Results, infants’ choices in Experiments 1–3. Age groups depicted separately. ‘p < .05.](image-url)
distinguish between some of the explanations for infants' choices in Experiment 1 raised above. If 8-month-olds generally privilege outcomes and utilize intentions only in the absence of conflicting outcome information, they should prefer the Failed Hinderer to the Failed Helper in Experiment 2. Alternatively, if 8-month-olds' choices resulted solely from evaluations of Pigs who successfully complete their intentions, then they should choose randomly in Experiment 2 (as every character fails). If 5-month-olds' evaluations are based on outcomes alone they should consistently choose the Failed Hinderer in Experiment 2, who is associated with a positive outcome; if 5-month-olds have difficulty with failed attempts or with conflicting information of any kind they should choose randomly between Failed Helper and the Failed Hinderer. As with Experiment 1, we supplemented infants' choice results with adults' interpretations/evaluations of videos of the same puppet events.

4. Experiment 2: do infants privilege mental states or outcomes in social evaluations?

4.1. Methods

4.1.1. Participants

In Experiment 2, sixteen 5-month-olds (11 boys; mean = 5.2) and sixteen 8-month-olds (8 boys; mean = 8.2) chose between a Failed Helper and a Failed Hinderer. Six additional 5-month-olds and 3 additional 8-month-olds participated but were excluded from analyses due to procedure error (2/0), fussiness (3/1), failure to choose either puppet (1/1) and family interference (0/1). 10 adults (5 males, 5 females) were also tested; 1 additional adult was tested but not included in the final sample as he was part of the group of 5 individuals who did not follow instructions.

4.1.2. Procedures

All procedures were identical to Experiment 1, except all participants observed and chose between a Failed Helper and a Failed Hinderer.

4.1.3. Counterbalancing and reliability

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Failed Helper (green/blue), (2) Failed Helper side during habituation (left/right of stage during box events), (3) Failed Helper event order (first/second), and (4) Failed Helper side during choice (left/right). A second independent coder unaware of the identity of the Pigs examined a random 30% of choices for reliability; the two coders reached 100% agreement. With adult participants, Failed Pig shirt color (green/blue) and order (first/second) was counterbalanced within each condition.

4.2. Results

4.2.1. Habituation

Eight-month-olds were shown an average of 8 habituation trials; 5-month-olds were shown 9; this number did
not differ by age ($t(30) = -1.0$, $p = .33$). One 8-month-old failed to reach the habituation criterion within 14 events. Attention to the first 6 habituation events did not differ within or across age or event type ($p$'s $> .23$).

4.2.2. Choice – infants

Infants' choices are depicted in Fig. 2. As in Experiment 1, 8-month-olds in Experiment 2 distinguished the Pigs, preferring the positively-intended Failed Helper to the negatively-intended Failed Hinderer (14/16 chose the Failed Helper, binomial $p < .005$). Five-month-olds chose randomly between the Pigs (6/16 chose the Failed Helper, $p = .45$); a significant effect of age (Fisher's Exact Test, $p < .005$). As in Experiment 1, there was no effect of attention during habituation on choice ($f(1,30) = .14$; $p = .71$; $\eta^2_p = .01$).

4.2.3. Choice – adults

Adults' written descriptions are listed in the Appendix, and their choices appear in Fig. 3. Nine of 10 adults “liked” the Failed Helper over the Failed Hinderer (binomial $p < .05$) and 9 of 10 identified the Failed Helper as “nicer” (binomial $p < .001$). Of the 9 subjects whose choices were consistent with infants’, 8 used the word “help” when describing the action of the Failed Helper.

4.3. Discussion

Eight-month-olds (and adults) in Experiment 2 preferred a positively-intended Failed Helper over a negatively-intended Failed Hinderer, even though the Protagonist experienced a negative outcome during Failed Helper events and a positive outcome during Failed Hinderer events. This finding suggests that, in addition to utilizing intention when outcomes match, 8-month-olds privilege intention over outcome in a situation involving failed attempts to help and harm. This rules out both of the alternative explanations for 8-month-olds' choices in Experiment 1 discussed above: 8-month-olds can both overcome potential interference from valenced Protagonist-outcomes (interference was present in every event of Experiment 2), and they can appropriately evaluate characters who fail to achieve their prosocial and antisocial goals (all Pigs failed in Experiment 2). Five-month-olds, on the other hand, chose equally between the Pigs in Experiment 2, failing to distinguish them based on either intention or outcome. This pattern of results tentatively suggests that 5-month-olds do not privilege outcome over intention in these puppet scenarios, but rather have difficulties with either (1) interpreting failed attempts in particular, or (2) processing conflicting information in general.

Yet the question remains: do infants at either age utilize Protagonist-outcome to evaluate others when only the outcomes differ, that is, when intention is equated and therefore uninformative? Though adults explicitly believe intention is relatively more important than outcome, outcome valence nevertheless plays a role in adults' punishment judgments (e.g., Cushman, 2008). Indeed, in the absence of relevant intentional information, it could be beneficial to use outcome-valence to inform one's social evaluations: those who are currently associated with good (bad) outcomes for others may be associated with good (bad) outcomes for you in the future. Alternatively, one might prefer individuals who are successful in carrying out their intentions, regardless of what outcomes they result in, as successful individuals are presumably more effective across the board. These possibilities were tested in Experiment 3, in which only the Protagonist’s outcome differed across puppets’ actions.

Infants in Experiment 3 saw Pigs who either both had positive or both had negative intentions, but who were differentially capable of achieving them, leading to distinct outcomes for the Protagonist: a Successful Helper (Pig intention to help = positive, Protagonist achieves goal = positive) and a Failed Helper (Pig intention to help = positive, Protagonist fails to achieve goal = negative; 3A), or a Successful Hinderer (Pig intention to hinder = negative, Protagonist fails to achieve goal = negative) and a Failed Hinderer (Pig intention to hinder = negative, Protagonist achieves goal = positive; 3B). Thus, infants chose between puppets who both intended to help or both intended to harm the Protagonist; the Protagonist experienced a good outcome with one pig and a bad outcome with the other.

5. Experiment 3A and 3B: do infants evaluate others based on outcomes when mental states are uninformative?

5.1. Methods

5.1.1. Participants

Sixteen 5-month-olds (10 boys, mean = 4;28) and sixteen 8-month-olds (9 boys; mean = 7;29) chose between a Successful Helper and a Failed Helper in Experiment 3A; sixteen 5-month-olds (10 boys; mean = 4;25) and sixteen 8-month-olds (8 boys; mean = 7;28) chose between a Successful Hinderer and a Failed Hinderer in Experiment 3B. Eight additional 5-month-olds and 6 additional 8-month-olds participated but were excluded from analyses due to procedure error (2/0), fussiness (1/1), failure to choose either puppet (3/4), inattentiveness (1/0), and family interference (1/1). 10 adults were tested in each of Experiment 3A and 3B (5 males, 5 females per condition). 1 additional adult was tested in 3A but not included in the final sample, as he was part of the group of 5 individuals who did not follow instructions.

5.1.2. Procedures

All procedures were identical to Experiment 1, except that infants/adults observed and chose between either a Successful Helper and a Failed Helper (3A) or a Successful Hinderer and a Failed Hinderer (3B).

5.1.3. Counterbalancing and reliability

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Failed Pig (green/blue), (2) Failed Pig side during habituation (left/right of stage during box events), (3) Failed event order (failure event first/successful event first), and (4) Failed Pig side during choice (left/right). A second independent coder unaware of the identity of the Pigs examined a
random 30% of choices for reliability; the two coders reached 100% agreement. With adult participants, Failed Pig shirt color (green/blue) and order (first/second) was counterbalanced within each condition.

5.2. Results

5.2.1. Habituation

In the Successful Helper versus Failed Helper condition, 8-month-olds were shown an average of 8 habituation trials; 5-month-olds were shown 11; this number differed by age (t(30) = −2.57; p < .05; η² = .18). Three 5-month-olds failed to reach the habituation criterion within 14 events. In the Successful Hinderer versus Failed Hinderer condition, 8-month-olds were shown an average of 10 habituation trials; 5-month-olds were shown 9; this number did not differ by age (t(30) = .67, p = .51; η² = .01). Two 8-month-olds and three 5-month-olds failed to habituate within 14 events. Attention did not differ across condition or event type, either within or across age (all p’s > .12).

5.2.2. Choice – infants

Infants’ choices are depicted in Fig. 2. Neither 5- nor 8-month-olds in Experiment 3 distinguished the characters in either condition. Infants chose equally between a Successful Helper and a Failed Helper (10 of 16 eight-month-olds and 5 of 16 five-month-olds chose the Successful Helper, binomial p’s > .10) and between a Successful Hinderer and a Failed Hinderer (7 of 16 eight-month-olds and 6 of 16 five-month-olds chose the Failed Hinderer, binomial p’s > .44). Across conditions, infants were no more likely to choose a Pig associated with a better Protagonist-outcome (17/32 eight-month-olds and 11/32 five-month-olds did so, Fisher’s Exact Tests, p’s > .10), nor were they more likely to choose Pigs who successfully completed their goals to help or to harm (19/32 eight-month-olds and 15/32 five-month-olds chose the Successful puppet, Fisher’s Exact p’s > .37).

Unlike in Experiments 1 and 2, choices in Experiment 3 did not differ by age (Fisher’s Exact Test, p = .21). As in Experiments 1 and 2, attention to the first 6 habituation events did not influence choice within or across age or condition (all Fs < 1; all p’s > .35). Finally, at each age it was possible to compare infants’ responses in Experiments 1 and 3. In particular, there was a (to adults) “more positive” character in each Experiment: in Experiment 1 the more positive character had the helpful intention; in Experiment 3 the more positive character was associated with the better outcome. For 8-month-olds, 28 of 32 infants chose the positively-intended character in Experiment 1, and 17 of 32 infants chose the character associated with the positive outcome in Experiment 3 (Fisher’s Exact p < .01). Unlike 8-month-olds, 5-month-olds’ pattern of choosing the more positive puppet did not differ by Experiment (p = 1).

5.2.3. Choice – adults

Adults’ descriptions are in the Appendix and their choices are depicted in Fig. 3. In Experiment 3A, 7 adults “liked” the Successful Helper and 3 “liked” the Failed Helper; this was not different from chance responding (binomial test with chance = .5, p = .34). Four adults identified the Successful Helper as “nicer,” the remaining 6 indicated that neither puppet was “nicer.” The rate of identifying the Successful Helper as nicer was no different from chance (binomial p = .86). All 10 subjects described the events as though both Pigs tried to help (see Appendix).

In Experiment 3B, 8 adults liked the Failed Hinderer, and 2 liked the Successful Hinderer (binomial p = .10). Despite this relatively high rate of preferring the Failed Hinderer, only 1 participant also identified the Failed Hinderer as nicer; the remaining 9 participants wrote that “neither” was nicest (binomial p < .001). All 10 subjects described the events as antagonistic between the pigs and the cow (see Appendix).

As with infants, it was possible to compare adults’ choices of the “more positive” character in Experiments 1 (intention) and 3 (outcome). This comparison revealed a significant effect of condition on who adults identified as “nicer” (2 × 3 Fisher’s Exact Test, p < .000001), but no effect of who adults “liked” (p = .41). Presumably, this discrepancy is due to adults’ relative willingness to differentially evaluate puppets based on outcome if there is nothing else to go on, but relative unwillingness to claim that puppets who have demonstrated the same intention are differentially nicer. This pattern of results highlights the critical role of intention in social evaluation in adults.

5.3. Discussion

In sum, Experiments 1–3 suggest that by 8 months of age, although not yet by 5 months, preverbal infants use intentions – but do not use outcomes – to evaluate others’ failed attempts to help and hinder third parties. Critically, these results are inconsistent with past research suggesting that young children focus mainly on outcomes (e.g., Piaget, 1932/1965), and support the possibility that young children fail to privilege intention in their social and moral judgments is due to methodological difficulties, not psychological ones. Adults’ responses were largely similar to 8-month-olds’, especially when adults were asked to identify which puppet (if any) was “nicer”. Although adults’ tendency to “like” characters associated with positive outcomes in Experiment 3 is consistent with research suggesting that adults judge others based on outcomes in some cases (e.g., Cushman, 2008; Nagel, 1979; Williams, 1981), adults’ relative unwillingness to identify those associated with positive outcomes as “nicer” suggests that liking those associated with positive intentions versus positive outcomes need not reflect the same evaluative processes. Indeed, adults largely described the puppet events in terms of helping/preventing language, suggesting they were interpreting the events as designed. Thus, although it is impossible to directly compare adults’ and infants’ interpretations of the same stimuli (as one cannot obtain a verbal description from preverbal infants) the consistency in 8-month-olds’ and adults’ responses supports the hypothesis that infants’ puppet choices reflect their social evaluation of those engaging in failed attempts to help and harm.

But the conclusion that 8-month-olds’ social evaluations truly privilege intention requires ruling out other, physical explanations for infants’ choices. For instance, inherent to this non-verbal experimental design was that...
Pigs demonstrated their intentions to help or to harm through distinct physical behaviors; specifically, Pigs who intended to help always engaged in coordinated action with the Protagonist and Pigs who intended to hinder always slammed the box lid; these physical acts occurred whether or not Pigs’ intentions were successful. Perhaps, then, 8-month-olds’ preferences in Experiments 1 and 2 did not result from intentional analyses at all, but instead from simple physical preferences such as liking coordinated action, disliking lid slamming, or both. Similarly, 8-month-olds’ failures to distinguish Pigs in Experiment 3 could be due to both Pigs exhibiting coordinated behavior with the Protagonist (3A), or both slamming the lid (3B), and not to a lack of intentional difference. Thus, in Experiment 4 eight-month-olds chose between Pigs who engaged in coordinated or slamming behaviors that were very similar to the behaviors of the Helpers/Hinderers, but were neither prosocial nor antisocial. Because 5-month-olds’ failed to distinguish coordinated versus slamming Pigs in any prior Experiment, only 8-month-olds were tested in Experiment 4. In addition, as adults’ descriptions of the events in Experiments 1–3 made no mention of the physical aspects of the displays (see Appendix), adults were not tested in Experiment 4.

6. Experiment 4: does a preference for coordinated-action over slamming account for 8-month-olds’ choices?

6.1. Methods

6.1.1. Participants

Sixteen 8-month-olds chose between a Coordinated and a Slamming puppet (8 boys; mean = 7.25). One additional 8-month-old was excluded due to family interference.

6.1.2. Procedures

6.1.2.1. Habituation. Events are depicted in Fig. 4. As in all previous experiments, at the start of each event the curtain raised to reveal the two Pigs at the rear of the stage and a clear box containing a toy at the center. In addition, a piece of black thread (against the black curtain this thread was not visible to the infants) was attached by a rubber band to the front of the box lid; the thread lead up and over the curtain behind the stage. Events began identically to previous experiments: the Protagonist enters, looks into the curtain behind the stage. Events began identically to previous experiments, at the start of each event the curtain tightened the thread attached to the lid of the box, leaving the lid ajar approximately 6 cm. First, the Protagonist jumped off the partially-open box and lay down next to the box (as in Successful Hinderer and Failed Helper events). Then, the Slammer moved forward and jumped onto the opposite corner of the lid, slammed it shut, and ran off-stage. Crucially, then, the physical behaviors of the Slammering character were the same as those of the Successful and Failed Hinderers in Experiments 1–3, but the slamming occurred only after the Protagonist had ‘given up’ (by putting his head down onstage) and was out of the way. Therefore, the slamming behavior did not itself block the Protagonist’s goal, nor did it demonstrate a negative intention to do so.

Coordination and Slamming events always resulted in the same negative Protagonist-outcome: the Protagonist failed to achieve his goal, and the trial ended with the Protagonist lying down next to the box. Because outcomes did not differ, this contrast was the most likely to reveal low-level physical influences on infants’ evaluations.

6.1.3. Counterbalancing and reliability

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Slamming Pig (green/blue), (2) Slamming Pig side during habituation (left/right of stage during box events), (3) Slamming Pig event order (first/second), and (4) Slamming Pig side during choice (left/right). A second independent coder unaware of the identity of the puppets examined a random 30% of choices for reliability; the two coders reached 100% agreement.

6.2. Results

6.2.1. Habituation

Infants were shown an average of 9 habituation trials; one infant failed to reach the habituation criterion within 14 events. Infants’ attention to the first 6 habituation events did not differ by event type ($F(1,15) = .03, p = .86; \eta_p^2 = .002$).

6.2.2. Choice

Infants’ choices are depicted in Fig. 5. Eight-month-olds in Experiment 4B did not distinguish between the Coordinated and Slamming puppets (7/16 chose the Coordinated puppet; binomial $p = .8$). As in previous studies, there was no affect of attention to habituation on infants’ choices ($F(1,15) = .06; p = .81; \eta_p^2 = .004$).

6.3. Discussion

Results from Experiment 4 bolster the claim that 8-month-olds’ preferences in Experiments 1–3 reflected intention differences rather than a preference for coordinated acts over slamming, and support the hypothesis that 8-month-olds engage in mentalistic evaluations of prosocial and antisocial individuals. Five-month-olds, on the other hand, did not distinguish characters in any condition in Experiments 1–3; the underlying structure of their social-evaluative capacities remains unclear. Although their failure to evaluate based on outcome in any condition
where they might have done so (Expts. 2 and 3) suggest that 5-month-olds do not privilege outcomes, this conclusion rests on an assumption not tested in Experiments 1–4: that 5-month-olds can distinguish characters who successfully carry out their prosocial and antisocial goals. Five-month-olds’ preference for Successful Helpers over Successful Hinderers in this box-opening scenario has previously been documented (e.g., Hamlin & Wynn, 2011); however, previous studies were carried out in a different laboratory and used different puppets. Therefore, Experiment 5 replicates this comparison. To gather more data about the validity of our puppet shows in tapping notions of helping and hindering, adults were also tested in Experiment 5.

7. Experiment 5: do 5-month-olds prefer Successful Helpers to Successful Hinderers?

7.1. Methods

7.1.1. Participants

In Experiment 5, sixteen 5-month-olds chose between a Successful Helper and a Successful Hinderer (9 boys; mean = 4.26). Seven additional 5-month-olds were excluded due to procedure error (4), and failure to choose either puppet (3). Ten adults (5 males, 5 females) were also tested; 2 additional adults were tested but not included in the final sample as they were part of the group of 5 individuals who did not follow instructions.

7.1.2. Procedures

All procedures were identical to Experiment 1, except infants/adults observed and chose between a Successful Helper and a Successful Hinderer.

7.1.3. Counterbalancing and reliability

The following were counterbalanced for infants in each age group and in each condition: (1) shirt color of the Successful Helper (green/blue), (2) Successful Helper side during habituation (left/right of stage during box events), (3) Successful Helper event order (first/second), and (4) Successful Helper side during choice (left/right). A second independent coder unaware of the identity of the puppets examined a random 30% of choices for reliability; the two coders reached 100% agreement. With adult participants, Successful Helper shirt color (green/blue) and order (first/second) were counterbalanced.

7.2. Results

7.2.1. Habituation

Five-month-olds were shown an average of 10 habituation trials. Two 5-month-olds failed to reach the habituation criterion within 14 events. Attention to the first 6 habituation events did not differ by event type ($F(1,15) = .19, p = .67; \eta^2_p = .01$).

7.2.2. Choice – infants

Choices are depicted in Fig. 5. Five-month-olds in Experiment 5 preferred a Successful Helper to a Successful Hinderer (13/16, binomial $p < .05$). There was no effect of attention to habituation on infants’ choices ($F(1,15) = .002; p = .97; \eta^2_p = .00$).
7.2.3. Choice – adults

Choices are depicted in Fig. 5. All 10 participants “liked” the Helper, and all 10 identified the Helper as “nicer.”

7.3. Discussion

Results from Experiment 5 replicate previous research suggesting that 5-month-olds evaluate third parties who successfully carry out their prosocial and antisocial intentions, leading to positive and negative outcomes for a Protagonist, respectively. This positive result supports the claim that 5-month-olds’ failure to distinguish the Pigs in Experiments 1–3 was due to difficulties with failed attempts in third party actions or to information conflict, rather than to an inability to evaluate puppets engaging in these particular social acts.

8. Summary and general discussion

8.1. Summary

The current studies examined the relative weight of actors’ mental states in preverbal infants’ evaluations of individuals who try but fail to help or harm third parties. Although previous work has demonstrated that young children base their explicit moral judgments primarily on outcome (Piaget, 1932/1965; see Karniol, 1978; Keasey, 1978; for reviews; see also Baird & Astington, 2004; Baird & Moses, 2001; Chandler et al., 2001; Killen et al., 2011; Zelazo et al., 1996), here 8-month-old preverbal infants’ social evaluations differed relative to puppet intention. Specifically, across Experiments 1 and 2, 8-month-olds preferred a puppet who tried to help a third party in its goal (and either succeeded or failed to help) over a puppet who tried to prevent the third party from achieving its goal (and either succeeded or failed to hinder), but did not distinguish puppets who had the same intentions to help or hinder but were associated with distinct outcomes for the recipient of helping and hindering. Five-month-olds, in contrast, did not consistently prefer any puppets in Experiments 1–3, neither distinguishing based on intention alone (Experiments 1 and 2) nor outcome alone (Experiment 3). Although it is difficult to interpret null results, it is noteworthy that 8-month-olds did not distinguish puppets based solely on Protagonist-outcome in Experiment 3, where intentions matched and only outcomes differed: 8-month-olds readily distinguished puppets in other conditions in which characters’ mental states differed (Experiments 1A, 1B, 2). This lack of preference tentatively suggests that privileging outcome is not a developmental precursor to privileging intention in infants’ evaluations of prosocial and antisocial others.

Although it is difficult to determine why 5-month-olds failed to distinguish the characters in Experiments 1–3, it is telling that both 5- and 8-month-olds failed to distinguish puppets solely on outcome in Experiment 3. That is, unlike what appears to be the case for explicit, verbal reports of intention(outcome tasks within and outside of the moral domain, the current studies did not reveal that a tendency to focus on outcome precedes a tendency to focus on intention in a spontaneous-response task. This pattern of results suggests that 5-month-olds may simply have trouble processing conflicting information of any kind, whether it be intentions versus outcomes, or something else entirely. Indeed, in a previous study of early social evaluations, 5-month-olds did not incorporate information about an actor’s previous behavior into their analyses of those who subsequently helped and hindered the actor, but 8-month-olds did (e.g., Hamlin et al., 2011). Critically, this study contained conflicting information (characters who had helped were subsequently harmed), but the conflict did not involve intention; yet, 5-month-olds failed. Thus, it seems possible that limitations in general information-processing abilities preclude 5-month-olds’ success on any complex social evaluation task. It is up to future study to better explore this possibility.

To determine whether adults view the puppet events as infants do, we replicated each experimental condition with adults and asked participants to provide a short written description of what they saw. Overall, results strongly indicated that adults viewed the events in a manner consistent with the current interpretation of infants’ choice data, describing the events as instances of helping and hindering and sometimes using evaluative language such as “nice” and “mean”. Adults’ choices demonstrate that they preferred positively-intended Pigs, even when Pig intention-valence opposed Protagonist outcome-valence, and identified the positively-intended pigs as nicer. Although in Experiment 3B adults tended to say that they “liked” the character associated with the better outcome, they nonetheless identified neither puppet as nicer, suggesting that intention and outcome evaluations need not reflect the same judgment criteria for adults.

8.2. An account of infants’ choices: implicit evaluations and spontaneous responses

How can one reconcile 8-month-olds’ choices, which seem to reflect an analysis and evaluation of the intentions of prosocial and antisocial others, with decades of evidence that young children privilege outcome in their moral judgments? While possible, it is not particularly parsimonious to posit that the tendency to privilege intention emerges by 8 months, and then social evaluation changes in early childhood to privilege outcome, only to change back to privileging intention in middle childhood. Nor is it satisfying to conclude that infants’ mentalistic judgments reflect some early evaluation system that is completely unrelated to later moral judgments: it is difficult to imagine a completely new system for moral evaluation developing when an alternative system for social evaluation is functioning reasonably well and could function as a building block for subsequent moral development.

Perhaps a more plausible possibility is one that is often raised to account for infants’ successes on spontaneous-response Theory of Mind tasks (e.g., Apperly, 2012; Baillargeon et al., 2010; Hutto, Herschbach, & Southgate, 2011 (and included papers); Leslie, German, & Polizzi, 2005; Senju, Southgate, White, & Frith, 2009) as well as spontaneous-response tasks within several other conceptual domains (e.g., Baillargeon, 2004; Carey, 1990; Gelman &
Baillargeon, 1983): Puppet choice may tap an implicit form of evaluation, and may be relatively unsusceptible to factors that hamper children on more explicit, elicited-response tasks. In particular, elicited-response intention-outcome tasks may be difficult for young children in that they require several skills in addition to representing actors' intentions and recipients' outcomes. Such requirements include simultaneously holding competing pieces of information in mind, selecting one piece of information to focus on, and inhibiting what is presumably a pre-potent tendency to respond based on salient real-world outcomes. Arguably, these skills could be considered domain-general executive function capacities rather than specific components of social or of moral evaluations, and executive functions are known to mature throughout infancy and beyond (e.g. Diamond, 1990; Zelazo et al., 2003). Thus, 8-month-olds may have succeeded in the current studies because the puppet-choice task was simply easy enough to allow them to demonstrate their (implicit) capacity for mentalistic evaluation.

8.3. Intention and outcome in failed attempts versus accidents

In addition to (apparently) being able to handle information conflict, 8-month-olds' success in the current studies may have been facilitated by the fact that, of all of the possible intention/outcome conflicts relevant to moral evaluations (failed attempts, accidents, false beliefs, etc.), it is perhaps easiest to interpret the mentalistic nature of failed attempts. In particular, whereas failed attempts unambiguously intend certain things (all of their behaviors are directed toward the intended goal), they are usually merely associated with positive and negative outcomes, rather than the cause of them. As all of the puppets' actions in these studies were directly relevant to their intention, perhaps this made the outcomes for the Protagonist easier to discount. In contrast, accidental outcomes are clearly caused whereas intentions are ambiguous or absent; perhaps outcomes are somewhat harder to discount in these cases. Others have suggested that the absence of saliently-caused outcomes makes it easier to interpret failed attempts than accidents mentalistically for adults and for children (e.g., Baird & Astington, 2004; Young & Saxe, 2009; see also Cushman et al., 2013; Moran et al., 2011). In addition, deficits with reasoning about failed attempts versus accidents have recently been distinguished neurologically in adults, suggesting they reflect discrete cognitive processes (e.g., Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young et al., 2010). Finally, to return to the review of infant mentalizing research from the introduction, there is considerable evidence that infants in the first year after birth utilize or privilege intention in their interpretation of others' failed attempts (e.g., Behne et al., 2005; Brandone & Wellman, 2009; Hamlin et al., 2009; Legerstee & Markova, 2007; Marsh et al., 2010; but see Ruffman et al., 2012). In contrast, to my knowledge the earliest evidence for infants' sensitivity to accidents comes later – at 14 months of age (Carpenter et al., 1998) – and there exists a clear asymmetry in the number of accident-mentalizing versus failed attempt-mentalizing papers in the literature (e.g., one versus many). Future work examining whether this distinction between failed attempts and accidents also exists in infants' social evaluations would help to shed light on the development of social evaluations in particular, as well as on the structure and development of mentalizing more generally.

Relatedly, 8-month-olds' failure to show outcome sensitivity (relative to their success at privileging intention) is something of a mystery. Not only do outcomes seem to be the primary focus of young children's explicit moral judgments, but also adults are sensitive to outcome when intentions are uninformative or even irrelevant (e.g., Cushman, 2008; Nagel, 1979; Williams, 1982). Infants' lack of sensitivity to outcome in the current studies has several possible explanations. First, as argued in the previous paragraph, it is possible that outcomes lack salience in failed attempts because they are not directly caused. Although we do not currently know whether infants can interpret accidental helping and hindering mentalistically, one could predict that they might also be sensitive to accidental outcomes, as accidental outcomes are caused directly and may be more salient. A second possibility is that infants are sensitive to outcomes in some cases of failed attempts, but that the outcomes in the current studies were simply not salient enough to inspire differential evaluation. Finally, it is at least possible that infants' social evaluations are truly insensitive to outcome. For instance, consider Piaget's claim (1932/1956) that young children's tendency to privilege outcome is the result of adult authority-figures imposing "moralities of constraint" on children, reacting and punishing based on rule-following and outcomes rather than intention-valence. As parents may not see infants as capable of acting autonomously, following rules, etc., they may not yet have imposed such constraints. As a result, infants' evaluations could be relatively more mentalistic than young children's. Indeed, though Piaget clearly considered explicit judgments to be most relevant to moral development, he himself noted frequently that young children possess an implicit grasp of the role of mental states in morality at some point before that grasp is reflected in their explicit judgments ("verbal thoughts"). As Carpenter (2009) writes: "it is not that [Piaget believed that] young children cannot reason autonomously; it is that they do not tend to do so in relationships of constraint" (pp. 282).

Piaget's observations, together with the current results, beg the question: how should one consider the role of implicit processes in development? Are implicit judgments irrelevant or less important than explicit ones? Of course, infants' evaluations are certainly limited, lacking much of what is involved in an adult system of moral reasoning. Adults reflect upon which kinds of actions are good and bad in different situations, hold normative notions of what individuals should do or must do, and routinely come together to explicitly set down these notions in the form of governments and religious institutions; preverbal infants are presumably incapable of such tendencies. Infants' choices may reflect little more than implicit positive feelings towards some individuals and implicit negative feelings toward others, which lead to approach/avoidance behaviors but little else. While this could be taken as reason to reject infants' evaluations as entirely irrelevant to
true moral judgments, I believe this would be imprudent: recent work suggests that many of adults’ moral judgments are themselves implicit and automatic in nature, with reasoning playing something of a supporting role (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001; Koenigs et al., 2007; Schnall, Benton, & Harvey, 2008; Valdesolo & DeSteno, 2006; Wheatley & Haidt, 2005; cf. Pizarro & Bloom, 2003). This research suggests that explicit reasoning may not be the most commonly-utilized mechanism by which the average adult reaches his or her moral judgments; if not, it becomes critically important to determine if and when infants’ implicit evaluations are based on similar criteria to adults’ explicit ones, as well as to determine when and how they differ.

8.4. Origins

Evidence for competence so early in human ontogeny naturally raises questions regarding how mentalistic evaluative abilities emerge. There are at least 4 (non-mutually-exclusive) possibilities, considered in turn. First and foremost is Piaget’s claim that subjective morality arises out of negotiating reciprocal and mutually-respectful relationships with peers. Although this possibility seems plausible to account for the emergence of explicit subjective morality during middle childhood, it seems highly unlikely to underlie the emergence of mentalistic social evaluation in infancy. Infants may regularly interact with other infants, but early peer interactions typically consist of “parallel play” (Parten, 1932) in which children independently pursue their own interests while sitting next to each other. While they do occasionally engage in rudimentary cooperative activities and even some conflict by their first birthday (see Hay, Caplan, & Nash, 2009, for review), these behaviors are quite uncommon before toddlerhood, and presumably absent in 5- to 8-month-olds. Furthermore, even if infants did sometimes experience cooperation and/or conflict with their peers before 8 months of age, it is not clear how a specific focus on mental states could emerge from these interactions: early relationships almost certainly lack the mentalistic social evaluation that occur within peer relationships in childhood.

A second possibility is that infants develop notions of helping and hindering via personal experience being helped and hindered. For instance, being fed could be perceived as helpful, and being prevented from playing with a desired object could be perceived as unhelpful; infants from birth presumably experience positive feelings when their goals are met and negative ones when they are blocked. Through experiencing these feelings, infants could come to form evaluations of both the actions and the individuals who perform them, and eventually extend these evaluations to the third party interactions they observe. But though this account seems plausible for how infants come to value successful helping over successful hindering, just how mentalistic social evaluations could come from these experiences is unclear. First of all, parents’ goals regarding their young infants are presumably almost always successful; but even if parents occasionally fail, how could infants learn that attempted but failed feeding (which leaves them feeling hungry) is actually better than attempted but failed object-taking (which leaves them playing with a desired object)? Indeed, action consequences that are personally experienced are probably extremely salient, resulting in less, rather than more, of a focus on mental states. Second, infants necessarily experience both helping and hindering from each of their primary caregivers; yet, their preferences suggest they readily distinguish helpful from harmful individuals. How could the capacity to distinguish individuals based on helping and hindering emerge through experience, as positive and negative behaviors are not linked to specific individuals in infants’ daily lives?

A third possibility is that, although physical limitations make it unlikely that 8-month-olds could yet engage in helping and hindering behaviors themselves, they develop these abilities via observing how adults react to others’ helpful and unhelpful acts in their daily lives. If infants regularly observe individuals react positively to helpful acts, and negatively to unhelpful acts, they might eventually glean from such reactions how one is supposed to act, and come to evaluate the behaviors of new individuals in new circumstances along these lines. Indeed, this observational mechanism might also account for infants’ emerging focus on mental states: although Piaget claimed that parents tend to focus on outcomes when distributing praise and blame (1932/1965), perhaps this is not the case in infancy, and infants in fact readily observe the primacy of mental states in their parents’ reactions. Although this possibility is promising, there is at least one prediction from this account that is not borne out in the current data (nor, to my knowledge, in any previous research of infant social evaluation). Specifically, infants with older siblings presumably have many opportunities to observe (or experience) the helpful and unhelpful behaviors of their brothers and sisters, and to observe how their caregivers react to these behaviors. First-borns, on the other hand, likely do not regularly have such experiences. Arguably, then, if observational experience is at the root of infants’ evaluative competence, then infants with older siblings should do better at social evaluation tasks than those without (or worse, if they observe parents imposing moralities of constraint on older siblings). There is no evidence for this in the current data at either age group: errors are distributed equally across first- and later-borns.

A fourth possibility, as raised in the introduction, is that the early emergence of mentalistic social evaluation reflects an unlearned capacity that supports group living and collective action (see Alexander, 1987; Cosmides & Tooby, 1992; deWaal, 2006; Henrich & Henrich, 2007; Joyce, 2006; Katz, 2000; Price et al., 2002). In particular, distinguishing those who may help you from those who may harm you is crucial to engaging in mutually-beneficial cooperative relationships, and those who are better able to distinguish helpers from harmers should reap the most benefits from social interaction. As mental states are arguably relatively better cues to one’s future behaviors than are outcomes s/he is currently associated with, cognitive biases to focus on mental states may have been selected for. Indeed, though it is unlikely that young infants themselves benefit extensively from the basic ability to distinguish friends from foes, and arguably are too
physically immature to profit from such an ability even if they had it, there are several examples of structures and capacities which emerge in development before they are specifically required (e.g., the sex organs; the addition/subtraction of large numbers (e.g., McCrink & Wynn, 2004); a preference for native-language speakers (e.g., Kinzler, Dupoux, & Spelke, 2007)). That said, it is conceivable that early social biases might inform several other developmental processes such as learning about novel objects in the environment; for example, infants might choose to emulate or to learn from the behaviors of certain individuals and not others (see, e.g., Hamlin & Wynn, 2012; Shutts, Kinzler, Houk, & Spelke, 2009). Perhaps, then, mentalistic social evaluation emerges early due to its adaptive benefits, both during long-term human cooperation and during short-term individual learning.

Of course, any core evaluative tendency will necessarily be influenced by the experiences described above (and many others) as children develop, and by the maturation of other cognitive functions. Many decades of previous research have been dedicated to elucidating the role of a variety of experiential and maturational processes in moral development (see Kagan & Lamb, 1987; Killen & Smetana, 2006 for extensive reviews), and this work need not be taken as fundamentally incompatible with classic moral development research. In particular, the influence of experience (even if large) or the existence of developmental change (even if extensive) does not in itself preclude the possibility that some prerequisites for an ability emerge early. The current results support this possibility for the socio-moral domain by documenting mentalistic evaluation of helpers and hinderers in the first year after birth; it will be up to future research to explore the nature of the link between infants’ mentalistic social evaluations and older children’s and adults’ moral ones.

In sum, the current studies suggest that several basic requirements for moral evaluation – evaluating others based on their third party prosocial/antisocial actions, distinguishing the mental states that motivate these actions from the outcomes they result in, and privileging mental states over outcomes during social evaluation – emerge by 8 months of age. These findings support claims that some foundational aspects of adult socio-moral judgments may be rooted in an early-developing capacity for third party social evaluation (e.g., Bloom, 2010; Bloom, 2012; Dunfield & Kuhlmeier, 2010; Hamlin, 2013; Hamlin et al., 2007; Olson & Spelke, 2008; Premack & Premack, 1994; Wynn, 2008).

**Acknowledgements**

Thank you to Janine Gellerman, Doan Le, Conor Steckler, and Wanying Zhao for their assistance and helpful comments, to other members of the Centre for Infant Cognition at UBC, to the parents and infants who participated, and to several anonymous reviewers for important advice regarding earlier versions of the manuscript. Art credit to Janine Gellerman. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada.

**Appendix A.**

Adults’ written responses in each condition. Adults were asked to “write a short sentence depicting what you saw” after they had seen both puppet events. Transcriber was blind to participant condition. Participant responses that were inconsistent with liking/identifying the “better” puppet as nicer are noted following the written description in Experiments 1, 2, and 5; all responses are given in Experiment 3 as they were highly variable.

Experiment 1A: Successful Helper versus Failed Hinderer

1. The pigs help the cow open the box. The first one pops it open, and the second helps the cow pry it open.
   a. Nicer = neither
2. When the cow tried to open the box, the pig in the blue shirt tried to stop him, and the pig in the green shirt helped him.
3. The cow plays with the box. The awesome pig tries to stop him but he can’t and the second pig helps him.
   a. Liked = Failed Hinderer
4. Cow tries to pull up lid. Pig in green shirt helps open lid. Pig in blue shirt jumps on the box, then cow opens the lid.
5. The lid of the box is hard to open. One pig helps by pulling open the lid with the cow. The other jumps on the lid to stop the cow, but then the cow opens it.
6. One pig opens the box lid for the cow. The other pig tries to slam it closed, but the cow gets inside.
7. The naughty pig tries to stop the cow from getting his toy, and the nice pig helps him open the box for his toy.
8. First time cow tries to open the box, pig in green shirt slams it, but cow opens it anyway. Second time, pig in blue shirt helps him open it.
9. The first pig jumps the box lid closed, but the cow gets his toy after he gets off. The second pig helps him open the box.
10. A pig puppet wearing a blue shirt helps a cow puppet open a box. Then a pig puppet wearing a green shirt tries to close the box lid, but the cow opens it after he gets off.

Experiment 1B: Failed Helper versus Successful Hinderer

1. The first pig can’t help, and the other jumps on the box so the cow can’t get inside.
2. As the cow puppet tried to open the box, the first puppet tried to help him get inside, then the second time the other pig slammed the box closed.

(continued on next page)
10. A pig tries to help the cow open the box, and the second pig tries to help.
9. Cow can't get in box, even when friend tries to help him.
8. A cow tries to open the lid of a box. The first pig jumps on the box lid. The cow is sad. The second pig tries to help him too, but can't. The second pig jumps on the box lid. The cow is sad.
7. Each time the cow tried to get into the box, he couldn't. The first time the pig stopped him, and the second time the other pig couldn't get the box open either.
6. Pig in green shirt thwarts cow. Pig in blue shirt tries to help; fails.
5. The puppet in the middle couldn't open the box lid, and the one on the right tried to help and the one on the left was just mean.
4. The puppets both prevent the cow from getting in the box, though the first one looked like he might be trying to help.
3. Stuffed animals play with a box together, and one of the puppets opens it. The other puppet tries to help and the other one on the left was just mean.
2. The pig pushed open the box lid, but the other one slammed it closed so the cow couldn't get inside.
1. The nice pig tries to help the cow open the box but can't, and the other one slammed the box closed but the cow got in anyway somehow.

Experiment 2: Failed Helper versus Failed Hinderer

1. The cow can't get in, and both try to help but only the green one can help.
3. Both pig puppets tried to be helpful but only the pig in the green shirt could open the box.
4. The first pig helps the cow open the box and the second pig can't help.
5. Both of the pigs help the cow but one is full of fail.
6. Two pig puppets watch a cow puppet try to open a box. Both of them attempt to assist him in opening the lid of the box. Only the pig in the blue shirt succeeds.
7. The show had a dude try and open a box and one guy couldn't help and the cow was sad and the other guy helped and the dude got the thing in the box.
8. One pig helps the cow open the box, the other doesn't.
9. The pig in the green shirt tries to help the cow open the box and the pig in the blue shirt helps him all the way.

Experiment 3A: Successful Helper versus Failed Helper

1. The cow can't get in, and both try to help but only the green one can help.
3. Both pig puppets tried to be helpful but only the pig in the green shirt could open the box.
4. The first pig helps the cow open the box and the second pig can't help.
5. Both of the pigs help the cow but one is full of fail.
6. Two pig puppets watch a cow puppet try to open a box. Both of them attempt to assist him in opening the lid of the box. Only the pig in the blue shirt succeeds.
7. The show had a dude try and open a box and one guy couldn't help and the cow was sad and the other guy helped and the dude got the thing in the box.
8. One pig helps the cow open the box, the other doesn't.
9. The pig in the green shirt tries to help the cow open the box and the pig in the blue shirt helps him all the way.

Experiment 3B: Failed Hinderer versus Successful Hinderer

1. The fist pig stops the cow from getting in the box, and the second tries to but gets off the box too soon so the cow gets inside.
2. The cow resists the pig in the blue shirt and gets in the box, but the pig in the green shirt wins at closing the box.
3. Both pig puppets tried to be helpful but only the pig in the green shirt could open the box.
4. The first pig helps the cow open the box and the second pig can't help.
5. Both of the pigs help the cow but one is full of fail.
6. Two pig puppets watch a cow puppet try to open a box. Both of them attempt to assist him in opening the lid of the box. Only the pig in the blue shirt succeeds.
7. The show had a dude try and open a box and one guy couldn't help and the cow was sad and the other guy helped and the dude got the thing in the box.
8. One pig helps the cow open the box, the other doesn't.
9. The pig in the green shirt tries to help the cow open the box and the pig in the blue shirt helps him all the way.
3. Two pigs bully the cow puppet by jumping on his box, but he gets inside the second time.
a. Liked = Failed Hinderer
b. Nicer = neither
4. The cow is trying to steal from the pigs and the pigs try to defend their toy.
a. Liked = Successful Hinderer
b. Nicer = neither
5. Both pig puppets are mean to the cow puppet. The first one stops him from getting the box open but the second doesn’t.
a. Liked = Failed Hinderer
b. Nicer = neither
6. Each pig is a jerk but the pig in the green shirt is less mean because he lets the cow in after all.
a. Liked = Failed Hinderer
b. Nicer = Failed Hinderer
7. First pig tries to stop the cow from getting in the box but the cow gets in. Second pig stops the cow.
a. Liked = Failed Hinderer
b. Nicer = neither
8. The guy in the green shirt got off the box too soon so the cow could get inside, and the guy in the blue shirt prevented the cow from getting inside.
a. Liked = Successful Hinderer
b. Nicer = neither
9. The puppet on the right made the cow get off of the box, and the puppet on the left tried to but the cow got in anyway.
a. Liked = Failed Hinderer
b. Nicer = neither
10. Two pigs and a cow fighting over a toy. The second pig in blue stops the cow from getting the toy.
a. Liked = Failed Hinderer
b. Nicer = neither

Experiment 5: Successful Helper versus Successful Hinderer

1. The pig in the green shirt opened the box lid for the cow, but the pig in the blue shirt jumped on the box and the cow laid down.
2. A pig wearing a green shirt helped a cow get a toy, and another pig wearing a blue shirt stopped a cow from getting a toy.
3. One pig was a jerk and another was nice.
4. The blue pig stopped the cow from getting the toy and the green pig helped him.
5. Cow tried to open box, one pig jumps on box, one pig helps.
6. When the cow tried to open the box, one of the pigs helped him lift the lid and the other jumped on the lid.
7. The cow pulled on the lid of the box, and the pig in the green shirt jumped on the box so the cow couldn’t open it, and the pig in the blue shirt opened the lid.
8. The cow tried to open the box. The blue shirted pig helped him open the box, and the green shirted pig shut the lid of the box.
9. The pig in the blue shirt helped the cow open the box. The pig in the green shirt shut the lid of the box.
10. The little cow puppet wanted the box open but the pig in the green shirt slammed it shut before he could open it, but then the pig in the blue shirt helped him.

Discarded subjects:
These subjects’ data was discarded due to failure to follow instructions. Their descriptions/choices are listed here.

1. (Expt 1B: Failed Helper versus Successful Hinderer) Hahaha the cow never gets in the box.
a. Liked = Successful Hinderer
b. Nicer = Failed Helper
2. (Expt 2: Failed Helper versus Failed Hinderer) Haha they all suck. (Note, response quite illegible, this was transcriber’s best guess).
a. Liked = Failed Hinderer
b. Nicer = Failed Helper
3. (Expt 3A: Successful Helper versus Failed Helper) Pigs and a cow puppet all try to open a box.
a. Liked = Successful Helper
b. Nicer = neither
4. (Expt 5: Successful Helper versus Successful Hinderer) The cow couldn’t open a box, and a pig helped him, and another was awesome and didn’t and the cow was all sad.
a. Liked = Successful Hinderer
b. Nicer = Successful Helper
5. (Expt 5: Successful Helper versus Successful Hinderer) There was a pig puppet who opened the box lid with a cow and another who was all “denied!”.
a. Liked = Successful Hinderer
b. Nicer = Successful Helper

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.04.004.

References


