Preschoolers’ social and moral judgments of third-party helpers and hinderers align with infants’ social evaluations

Julia W. Van de Vondervoort*, J. Kiley Hamlin

Department of Psychology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

**A R T I C L E   I N F O**

Article history:
Received 1 November 2016
Revised 5 June 2017
Available online 17 August 2017

Keywords:
Children
Social judgments
Moral judgments
Moral development
Social cognition
Evaluations
Explanations

**A B S T R A C T**

Two experiments explored preschoolers’ social preferences and moral judgments of prosocial and antisocial others. In Experiment 1, 3- to 5-year-olds \((N = 74)\) observed helping and hindering scenarios previously used to explore sociomoral evaluation in preverbal infants. Whereas 3-year-olds in Experiment 1 did not reliably distinguish between the helper and hinderer when reporting social preferences or moral judgments, both 4- and 5-year-olds preferred the helper, judged the helper to be “nicer” than the hinderer, selectively allocated punishment to the hinderer, and were able to justify their punishment allocations. A simplified procedure and the addition of comprehension questions in Experiment 2 \((N = 24)\) improved 3-year-olds’ performance, suggestive that their performance in Experiment 1 was likely due to processing or memory difficulties rather than an inability to engage in explicit social and moral evaluation. These studies reveal that young children readily interpret helping and hindering scenarios as socially and morally relevant.

© 2017 Elsevier Inc. All rights reserved.

**Introduction**

Humans are remarkably judgmental. One example of this is our readiness to evaluate individuals based on their prosocial and antisocial actions; we condemn those who willfully harm and see them...
as deserving of punishment, and we celebrate those who selflessly help and see them as deserving of praise. Notably, humans make these judgments even in third-party situations when we are not personally invested in the welfare of those who are harmed and helped. These third-party moral judgments are commonplace in adult life across cultures (Brown, 1991), and a growing body of work suggests that such judgments are also evident during childhood.

Young children make moral judgments, suggestive that a moral sense exists (but is immature) during early childhood. Specifically, research has demonstrated that young children's explicit moral judgments are sensitive to some of the same factors that influence adults' moral judgments. For example, under certain testing conditions, 3-year-olds' explicit judgments are sensitive to the role of intentions in assigning moral praise and blame (e.g., Nelson, 1980; Nobes, Panagiotaki, & Pawson, 2009; Yuill & Perner, 1988). In addition, young children are sensitive to the difference between moral and nonmoral concerns; for example, 3-year-olds judge moral transgressions (e.g., hitting another child) to be more generalizable wrong than social conventional transgressions (e.g., not saying “please”). By 3 or 4 years of age, children also judge moral transgressions to be more serious than conventional transgressions, more independent of explicit rules, and wrong regardless of whether the transgression is sanctioned by an authority figure (e.g., Smetana & Braeges, 1990). These distinctions are reflected in children's verbal justifications of their judgments; whereas moral justifications tend to focus on issues of welfare and fairness, conventional justifications focus on concerns about authority, social order, and rules (Nucci & Weber, 1995; see Smetana, 2006, and Smetana, Jambon, & Ball, 2014, for reviews). Finally, by 4 years of age, children treat moral versus immoral others differently (e.g., distributing more treats to prosocial vs. antisocial others) and explicitly justify nonegalitarian treatment in terms of individuals' morally relevant behaviors (Kenward & Dahl, 2011).

Other work using implicit measures has also demonstrated that young children appreciate the nature of moral concerns. For example, 3-year-olds spontaneously protest when others are the victims of moral transgressions (Rossano, Rakoczy, & Tomasello, 2011; Vaish, Missana, & Tomasello, 2011) and protest third-party moral (but not conventional) transgressions regardless of perpetrators' group membership (Schmidt, Rakoczy, & Tomasello, 2012). These results suggest that children are sensitive to the generalizability of moral principles. In addition, 3-year-olds selectively avoid directing prosocial behaviors toward those who have previously harmed (or merely tried to harm) others (Vaish, Carpenter, & Tomasello, 2010; see also Dahl, Schuck, & Campos, 2013; Hamlin, Wynn, Bloom, & Mahajan, 2011). Altogether, these studies demonstrate that children both explicitly and implicitly evaluate morally relevant actions by 3 years of age.

Although the studies reviewed above suggest that the moral sense is operational fairly early in life, measures that assess children's verbal and complex behavioral responding necessarily restrict the exploration of early moral development to children who can reliably generate these responses. Therefore, during recent years researchers have developed alternative nonverbal methodologies to explore whether very young children are sensitive to third-party morally relevant action (see Hamlin, 2013b, for a review). In these studies, preverbal infants are shown puppet shows featuring a puppet protagonist who is unable to accomplish a goal—for example, trying but failing to climb a hill, open a box, or play with a ball. The protagonist’s goal is facilitated by a “helper” puppet who pushes the protagonist up the hill, helps him open the box, or returns the dropped ball; the protagonist’s goal is blocked by a “hinderer” puppet, who pushes the protagonist down the hill, slams the box closed, or takes the ball away. Infants are then presented with the helper and hinderer side by side, and their selective looking and reaching behaviors are taken as evidence of a preference for one puppet or the other.

Results from studies using these methods have demonstrated that infants prefer helpers over hinderers from soon after birth. By just 3 months of age, infants look longer at individuals who help a protagonist achieve a goal compared with individuals who thwart the protagonist’s efforts (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2010). Once infants can make visually guided reaches, they selectively reach for helpers over hinderers (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007; Scola, Holvoet, Arciszewski, & Picard, 2015; but see Salvadori et al., 2015). Critically, these preferences appear to reflect social concerns rather than merely perceptual concerns; infants show no preference for “helpers” who direct similar physical behaviors toward inanimate objects (Hamlin & Wynn, 2011; Hamlin et al., 2007, 2010) or toward an animate character without a clear unfulfilled goal (Hamlin, 2015; see also Scarf, Imuta, Colombo, & Hayne, 2012).
These studies suggest that infants’ preferences are social; however, they do not specify the nature of the social preference. Indeed, infants’ choices may reflect nonmoral social concerns, including analyses of who might belong to an infant’s social group and who might benefit the infant in the future (e.g., Baillargeon et al., 2015; see also Tafreshi, Thompson, & Racine, 2014; but see responses by Hamlin, 2014a, and Dahl, 2014). Potentially consistent with this social but not moral interpretation, infants’ preferences for helpers versus hinderers are influenced by the protagonist’s similarity to infants themselves (Hamlin, Mahajan, Liberman, & Wynn, 2013); infants prefer those who help similar others but also prefer those who hinder dissimilar others. Relatedly, from 4.5 months of age, infants prefer those who help previously prosocial others but also prefer those who harm previously antisocial others (Hamlin, 2014b; Hamlin et al., 2011). Although these judgments could reflect a sense of who deserves reward and punishment (presumably a moral concern), they could instead reflect affiliative judgments, whereby infants think “the enemy of my enemy is my friend” (e.g., Heider, 1958).

On the other hand, the preferences described above may be based in moral concerns. Indeed, infants may select hinderers of hinderers because they feel that antisocial others should be punished (see also Meristo & Surian, 2013). Furthermore, certain conceptions of morality highlight the role of group-level concerns (e.g., Haidt & Joseph, 2008; Schweder, Much, Mahapatra, & Park, 1997), and so the discovery that infants’ preferences are sensitive to target similarity does not necessarily indicate that they are not of a moral nature. In addition, other studies have shown that infants’ evaluations are sensitive to factors that influence adults’ moral judgments, for example, whether individuals tried to help or hinder even if they did not succeed (Hamlin, 2013a; see also Lee, Yun, Kim, & Song, 2015; Vaish et al., 2010) and whether individuals knew they were helping or hindering when they acted (Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013; see also Choi & Luo, 2015; Meristo & Surian, 2013; Sloane, Baillargeon, & Premack, 2012). Finally, in a recent study that directly pit self-interest against moral value, infants incurred costs (i.e., accepted fewer graham cracker treats) to avoid interacting with a hinderer, suggestive that infants can be motivated by more than simple self-interest (Tasimi & Wynn, 2016).

In sum, the body of work described above demonstrates that infants’ preferences in helping and hindering puppet scenarios are influenced by some of the same factors that influence adults’ social and moral judgments. Unfortunately, because preverbal infants can neither respond selectively following different prompts nor explain their responses, it is difficult to determine whether infants’ responses to helpers and hinderers are based on social or moral concerns. Conversely, it is possible to distinguish between social and moral responding in older children, who are more linguistically advanced than infants. Specifically, one can ask who children prefer (a positive social judgment), who they think is nice (a positive moral evaluation), and who they think should be punished (allocating a negative moral consequence). To our knowledge, there is limited work exploring young children’s social and moral responses to third-party helping and hindering scenarios designed for infants (but see Kenward & Dahl, 2011, for preschoolers’ judgments following a show not previously tested on infants, and see Buon et al., 2014, for a comparison between infants’ and toddlers’ responses to a prosocial and violent antisocial other); the current experiments aimed to fill this gap.

The current experiments explored children’s explicit reasoning about characters involved in live helping and hindering puppet shows. Here, 3- to 5-year-olds were presented with box and ball helping and hindering puppet shows originally used in Hamlin and Wynn (2011). These shows were chosen due to their use in many publications to date (e.g., Hamlin, 2013a, 2014b; Hamlin et al., 2011, 2013; Salvadori et al., 2015; see also Scola et al., 2015) and because their use of simple props made them ideal for transporting to day-care centers, where most participants were tested. After being presented with the puppet scenarios, each child was asked to report (a) who they “like,” (b) who is “nicer,” and (c) who “should get in trouble.” Following children’s identification of who should get in trouble, children were asked to explain their judgment. These questions were adapted from previous work in the moral development literature (e.g., Baird & Astington, 2004; Costanzo, Coie, Grumet, & Farnill, 1973; Cushman, Sheketoff, Wharton, & Carey, 2013; Leslie, Mallon, & DiCорcia, 2006; Zelazo, Helwig, & Lau, 1996) and were chosen because they reflect similar methods to those previously used with infants (e.g., asking “who do you like?” while presenting the helpful and unhelpful puppets; see Hamlin, 2013b).
We predicted that at all ages children’s social preferences and positive moral evaluations would favor helpers over hinderers but that children would select hinderers over helpers as recipients of punishment. That said, there are two alternative possibilities: (a) that children would interpret the puppet shows as social but not moral and so would reliably prefer helpers over hinderers but not identify helpers as nicer or hinderers as deserving of punishment and (b) that children would readily identify the moral value of helpers and hinderers but that their social judgments would reflect more idiosyncratic responding, resulting in chance-level responding on the liking question. Based on previous work showing that 3-year-olds can have difficulty in producing interpretable responses to open-ended questions (e.g., Kenward & Dahl, 2011; Killen, Breton, Ferguson, & Handler, 1994), we assumed that older children would provide more coherent verbal justifications than 3-year-olds. In addition, we predicted that older children would be more likely than 3-year-olds to reference relevant social and/or moral considerations as the reason for their punishment allocations.

**Experiment 1**

**Method**

**Participants**

Children in both experiments were recruited from a mid-size North American city through hospitals and day-care centers. In total, 25 3-year-olds ($M_{age} = 3;6$ years;months, range = 3;0–3;11; 13 girls), 24 4-year-olds ($M_{age} = 4;5$, range = 4;0–4;11; 11 girls), and 25 5-year-olds ($M_{age} = 5;4$, range = 5;0–5;9; 13 girls) were tested in a university research center or the children’s day-care centers (the preset sample size was 24 children per condition; an extra 3-year-old and an extra 5-year-old were run due to scheduling issues). An additional 6 3-year-olds, 4 4-year-olds, and 2 5-year-olds were seen but replaced due to unwillingness to watch the puppet show and/or to point to a puppet in response to questions ($n = 5$), interference from parents or teachers ($n = 2$), procedure errors ($n = 4$), or failure to complete a warm-up task that required English comprehension ($n = 1$).

**Procedure**

Children first completed a warm-up task with an image of a playground. Children were asked to point to a swing and a slide, to verbally identify the color of a toy, and to identify their favorite outside activity. As noted above, 1 child was removed from the sample due to an inability to locate the swing or slide; children were retained in the sample if they pointed to these items (i.e., verbal responses were not required).

Children then watched live puppet shows featuring a protagonist struggling to achieve his goal to open a box or retrieve a dropped ball in the presence of two other characters. All shows were identical to those used in previous infant studies (e.g., Hamlin & Wynn, 2011) except for two factors. First, in infant studies puppet shows are performed at one end of a long table, with a curtain lowering to occlude the “stage” between events; puppeteers are kept out of sight behind an additional curtain at the back of the stage. In contrast, in the current study shows were enacted either on the floor or on a table directly in front of children and with the puppeteer fully visible because this was required to test in day-care centers. Second, a few nonvalenced words were added to the shows for narration because pilot testing revealed that this helped children to remain attentive to the events. Each child watched both the box and ball scenarios in counterbalanced order.

**Box scenario.** Three characters performed the box puppet show. The experimenter enacted the protagonist saying “Hello!”, walking up to a clear box containing a purple whale toy, looking through the side of the box while saying “Look, a toy!”, and then unsuccessfully attempting to open the box five times. On the third attempt, the protagonist said, “Too heavy!” On the fifth attempt, the helper assisted the protagonist in opening the box while saying “Open!” or the hinderer prevented the protagonist from opening the box by slamming the lid shut while saying “Close!” All narrations in both scenarios were spoken in a high-pitched positive voice to indicate that it was the puppet speaking rather than the experimenter; speech was not modulated based on the valence of the behavior.
occurring. Children were shown two helping events and two hindering events in alternation for a total of four events.

**Ball scenario.** A different set of three characters performed the ball puppet show. The experimenter enacted the protagonist saying “Hello!”, walking up to a ball and saying “Look, a ball!”, and then bouncing the ball three times. Following the third bounce, the protagonist dropped the ball to one side while saying “Whoops!” The ball was either picked up and returned to the protagonist by the helper (who said “Here!” while rolling the ball to the protagonist) or taken away by the hinderer (who said “Good-bye!” while running away with the ball). Children saw two helping events and two hindering events in alternation for a total of four events.

**Social preference and moral judgments.** After watching each of the above scenarios, children were presented with the helper and hinderer and asked (in counterbalanced order) which puppet they preferred (i.e., “Which one of these guys do you like the most?”) and which puppet was nicer (i.e., “Which one of these guys was nicer?”). After children indicated a positive moral evaluation, they were asked whether the selected puppet was a “little bit nice or a lot nice” (order counterbalanced; we planned to examine niceness judgments on a 3-point scale from not nice to a lot nice, but because most children at each age responded that the selected puppet was a lot nice regardless of whether they selected the helper or hinderer, this question is not considered further). To reduce the likelihood that children would select the same puppet across the liking and nicer questions due to reaching perseveration, children were asked to point to each puppet in between these questions (e.g., “Point to the guy with a red shirt. Right! Point to the guy with the green shirt. Right!”; all children correctly identified the puppets’ shirt color, and note that no reinforcement was given following responses to the liking and nicer questions). Children were then asked to allocate punishment to one of the puppets (i.e., “I think that one of these guys should get in trouble. Who should get in trouble?”) and then to justify this choice (i.e., “Why should he get in trouble?”). If children did not provide a verbal response to the last question, they were prompted (e.g., “What do you think?”). Children were asked to provide only one verbal response to prevent interference between explanations about the same puppet show.

Two sets of three puppets were used to complete the two scenarios: a duck (protagonist) and two rabbits wearing red and green shirts (helper and hinderer; identity counterbalanced) and a wolf (protagonist) and two moose wearing blue and yellow shirts (helper and hinderer; identity counterbalanced). The duck and rabbits were used in the first round, and the wolf and moose were used in the second round. Because the order of scenarios was counterbalanced across children, which set of puppets enacted which scenario (box or ball) was also counterbalanced. Additional counterbalanced variables were the order of the events within each show (helping first or hindering first) and the side of the helper and hinderer puppets (right or left). For the question period, the helper and hinderer remained on the same side as during the show.

**Transcription and coding procedure**

When permitted by children’s parents and feasible within the day-care setting, participation in the study was audio and visually recorded. A research assistant transcribed children’s explanations regarding why the selected puppet should get in trouble from these recordings. When recording was not permitted (11 of 74 children), children’s explanations were transcribed during the study by the experimenter. Two additional independent research assistants who had not assisted with data collection or transcription then coded the content of children’s explanations according to the categories detailed below.

Based on an initial review of children’s explanations by the authors, it was determined that children’s responses were either informative, meaning the children referred to something about the puppet or puppet show, or uninformative, meaning they did not.

**Uninformative responses.** Uninformative responses were highly variable and included cases in which children did not provide a verbal response, provided an unintelligible verbal response, or provided a verbal response that did not include a justification for their selection. These verbal responses included statements entirely unrelated to the characters or events in the puppet show.
(e.g., “I don’t like skating”), repetitions of the question (e.g., “Because he should get in trouble”), statements not providing a reason (e.g., “because”), and statements implying that the children were unsure (e.g., “I don’t know”).

**Informative responses.** Informative responses were those that related to the characters from the show and/or an action performed and took one of the following forms:

- **Relevant action**—Children appealed to the performance of the hindering action from the box or ball scenario (e.g., “because he closed the box,” “because he didn’t give the ball back”).
- **Irrelevant action**—Children appealed to the performance of a negative action that had not occurred in the box or ball scenario (e.g., “because he hit somebody”). These responses might indicate that, when asked why a puppet should get in trouble, some children simply generated actions that generally lead someone to punishment.
- **Relevant valence**—Children appealed to the positive or negative valence of the puppet or his actions; these statements were directly related to the puppet shows (e.g., “because he is mean [pointing to selected puppet],” “because this guy is nicer” [pointing to unselected puppet]).
- **Irrelevant valence**—Children appealed to the positive or negative valence of the puppet or his actions; these statements were not directly related to the puppet shows (e.g., “because he’s happy”).
- **Nonsocial considerations**—Children’s responses did not reference positive or negative actions or valence. Responses in this category included physical descriptions of the puppet (e.g., “because he has long ears”), general disliking of the selected puppet (e.g., “he’s not the most favorite thing”), references to the previous puppet show (e.g., “because he closed it” following the ball show), and ambiguous statements about the puppets or shows (e.g., “it wasn’t saying anything”).

Each child’s explanation was coded by two independent research assistants for the presence or absence of each response type; coders were blind to the referent (helper or hinderer) of the explanation but not to whether the child was responding after the box or ball show. Informative response types were not mutually exclusive. Reliability across the six categories was strong (average Cohen’s kappa = .864, range = .701–1.000; see McHugh, 2012). Disagreements (which involved 14 of 148 responses) were then reviewed by the first author and the two coders. Of these 14 disagreements, 11 were immediately recognized as mistakes by one of the coders, typically due to failure to consider whether the child was responding to a box show or ball show (e.g., “because he closed it” following the ball show). The other 3 statements were resolved by discussion among the two coders in the presence of the first author.

**Results**

**Social preference and moral judgments**

**Confirmatory analyses.** For each test question (liking, niceness, or trouble), children received a score of 1 if they responded in the direction of the hypothesis (i.e., liked the helper, judged the helper as nicer, or assigned punishment to the hinderer) or 0 if they responded against the direction of the hypothesis. Children’s scores were summed across the two rounds, resulting in three scores between 0 and 2 per child, one for each question type (see Fig. 1). We predicted that, at all ages, children’s social preferences and positive moral evaluations would favor helpers over hinderers but that children would select hinderers over helpers as recipients of punishment. To determine at which ages children did respond as predicted, a series of one-sample t tests were conducted comparing children’s liking score, niceness score, and trouble score at each age with a chance score of 1.

---

1 One 5-year-old in Experiment 1 pointed to the helper and appealed to the helping action from the box scenario when explaining why he should get in trouble (i.e., “because he opened it”).

2 Explanation coding was originally conducted by two different research assistants and resulted in substantially lower reliability (average Cohen’s kappa = .791, range = .340–1.000). The coding scheme was then clarified, a more extensive training procedure was created, and the data were recoded by two different research assistants.
Contrary to our predictions, 3-year-olds did not reliably distinguish between the helper and hinderer when providing social or moral judgments. The 3-year-olds’ liking score \( (M = 1.080, SD = 0.640) \) and niceness score \( (M = 1.200, SD = 0.707) \) did not differ from chance, one-sample \( t \) tests comparing each score between 0 and 2 with a chance score of 1. \( t(24)_{\text{liking}} = 0.625, p = .538, d = 0.125; t(24)_{\text{niceness}} = 1.414, p = .170, d = 0.283 \). On the other hand, 3-year-olds reliably identified the hinderer as the appropriate recipient of punishment \( (M = 1.320, SD = 0.748) \), one-sample \( t \) test, \( t(24) = 2.138, p = .043, d = 0.428 \). In contrast to 3-year-olds, both 4- and 5-year-olds selectively preferred the helper, judged the helper to be nicer, and allocated punishment to the hinderer. The 4-year-olds’ liking score \( (M = 1.417, SD = 0.830) \), niceness score \( (M = 1.542, SD = 0.721) \), and trouble score \( (M = 1.583, SD = 0.717) \) all exceeded chance in the direction of the hypothesis, one-sample \( t \) tests, \( t(23)_{\text{liking}} = 2.460, p = .022, d = 0.502; t(23)_{\text{niceness}} = 3.680, p = .001, d = 0.751; t(23)_{\text{trouble}} = 3.984, p = .001, d = 0.813 \). Likewise, 5-year-olds’ liking score \( (M = 1.640, SD = 0.700) \), niceness score \( (M = 1.880, SD = 0.332) \), and trouble score \( (M = 1.880, SD = 0.332) \) all exceeded chance in the predicted direction, one-sample \( t \) tests, \( t(24)_{\text{liking}} = 4.571, p < .001, d = 0.914; t(24)_{\text{niceness}} = 13.266, p < .001, d = 2.653; t(24)_{\text{trouble}} = 13.266, p < .001, d = 2.653 \). See Table 1 in the online supplementary material for a summary of individual patterns of responding within each round.

Exploratory analyses. Although differences in performance across age were not predicted, the analyses reported above reveal that whereas older children reliably formed both social and moral evaluations favoring the helper, 3-year-olds only reliably distributed punishment to the hinderer. To examine how age and/or question type influenced children’s tendency to respond in the direction of the hypothesis, we conducted a mixed-effects analysis of variance (ANOVA) with liking, niceness, and trouble scores as within-participants variables (repeated measures) and children’s age in years (3, 4, or 5) and children’s gender (female or male) as between-participants factors. This analysis revealed a main effect of question type, \( F(1.860, 126.478) = 5.170, p = .008, \eta^2_p = .071 \), and a main effect of age, \( F(2, 68) = 7.309, p = .001, \eta^2_p = .177 \). There was no interaction between question type and age, \( F(3.720, 126.478) = 0.220, p = .916, \eta^2_p = 0.006 \), and no main effect or interactions involving gender (all \( Fs < 2.420, all ps > .124 \)).

To probe the main effect of question type, a series of paired-samples \( t \) tests using the Bonferroni-corrected alpha value of .017 (.05/3) were used to compare scores on each question type across age. These tests revealed that children were more likely to respond in the direction of the hypothesis when asked which puppet should get to trouble (trouble scores: \( M = 1.595, SD = 0.660 \)) compared with which puppet they prefer (liking scores: \( M = 1.378, SD = 0.753 \)), paired-samples \( t \) test, \( t(73) = 2.704, p = .009, d = 0.314 \). There was no difference between liking scores and niceness scores (\( M = 1.541, SD = 0.666 \)), paired-samples \( t \) test, \( t(73) = 2.250, p = .027, d = 0.262 \), or between niceness scores and trouble scores, paired-samples \( t \) test, \( t(73) = 1.000, p = .321, d = 0.116 \). To explore the main effect of age, a series of independent-samples \( t \) tests using the Bonferroni-corrected alpha value of .017 (.05/3) were used to compare overall scores (between 0 and 6) across the three question types at each age. These tests revealed that 5-year-olds (\( M = 5.400, SD = 1.258 \)) were more likely to respond in the direction of the hypothesis across test questions compared with 3-year-olds (\( M = 3.600, SD = 1.683 \)), independent-samples \( t \) test, \( t(48) = 4.282, p < .001, d = 1.236 \). There was no difference between 3-year-olds’ overall scores and those of 4-year-olds (\( M = 4.542, SD = 2.000 \)), independent-samples \( t \) test, \( t(47) = 1.786, p = .081, d = 0.521 \), or between 4-year-olds’ scores and 5-year-olds’ scores, independent-samples \( t \) test, \( t(38.482) = 1.790, p = .081, d = 0.577 \).

3 Mauchly’s test of sphericity indicated that the assumption of sphericity had been violated (\( \chi^2 = 13.370, p = .001 \)). Because the estimated epsilon was greater than .750 (Greenhouse–Geisser: \( \varepsilon = .847 \); Huynh–Feldt: \( \varepsilon = .930 \)), we applied the Huynh–Feldt correction (see Girden, 1992).

4 We note that analyzing age continuously (in months) rather than categorically (in years) suggests the same pattern of results. Specifically, a series of linear regression analyses examined the effect of age in months and child’s gender (female or male) on liking, niceness, and trouble scores. Comparing the results of each regression analyses with the Bonferroni-corrected alpha value of .017 (.05/3) indicated an effect of age in months and gender on liking scores, \( F(2, 71) = 5.785, p = .005, R^2 = .140 \), nicer scores, \( F(2, 71) = 11.067, p < .001, R^2 = .238 \), and trouble scores, \( F(2, 71) = 6.126, p = .004, R^2 = .147 \). In each case, this effect was due to a positive relation between age in months and scores (liking scores: \( t = 3.138, p = .002, B = .027 \);niceness scores: \( t = 4.433, p < .001, B = .031 \); trouble scores: \( t = 5.124, p = .001, B = .053 \)); there was no independent effect of gender (liking scores: \( t = 1.495, p = .139, B = -.247 \); nicer scores: \( t = 1.834, p = .071, B = -.252 \); trouble scores: \( t = 1.320, p = .191, B = -.190 \)).
Explanations regarding punishment

Confirmatory analyses. To ensure that each child contributed equally to the dataset and to avoid over-representing talkative children (who may have provided several types of informative responses), instances of each explanation type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, and nonsocial considerations) are represented as proportions and averaged across the two rounds (see Table 1). We predicted that younger children would provide less interpretable explanations than older children. To test this prediction, we used a factorial ANOVA to examine the effect of children’s age in years (3, 4, or 5) and children’s gender (female or male) on the proportion of uninformative responses across rounds. This test revealed that the proportion of uninformative responses differed among the ages tested, $F(2, 68) = 10.182, p < .001, \eta^2_p = .230$; there was no main effect of gender or interaction between age and gender (both $F$s < 1.363, all $p$s > .262). To determine the nature of this difference, a series of independent-samples $t$ tests compared the proportion of uninformative responses across ages. These analyses revealed that 3-year-olds ($M = .520, SD = .510$) were more likely to provide uninformative responses than 4-year-olds ($M = .167, SD = .381$), independent-samples $t$ test, $t(44.357) = 2.756, p = .008, d = 0.828$, or 5-year-olds ($M = .040, SD = .200$), independent-samples $t$ test, $t(31.214) = 4.382, p < .001, d = 1.569$. This is consistent with previous work showing that it can be difficult to elicit coherent justifications from 3-year-olds (Kenward & Dahl, 2011; Killen et al., 1994). The 4- and 5-year-olds were equally (un)likely to provide uninformative responses when explaining their allocations of punishment, independent-samples $t$ test, $t(34.483) = 1.449, p = .156, d = 0.494$.

We also predicted that older children would provide more relevant references to the puppet’s actions or valence than younger children. To test this prediction, we combined the appeals to relevant actions and relevant valence into a single proportion of relevant responses. A factorial ANOVA examining the effect of children’s age in years (3, 4, or 5) and children’s gender (female or male) on the proportion of relevant responses revealed that the proportion of relevant responses differed among the ages tested, $F(2, 68) = 11.075, p < .001, \eta^2_p = .246$; there was no main effect of gender or interaction between age and gender (both $F$s < .752, all $p$s > .475). To determine the nature of this difference, a series of independent-samples $t$ tests compared the proportion of relevant responses across ages. These analysis revealed that 3-year-olds ($M = .280, SD = .458$) were less likely to appeal to relevant actions and valence than 4-year-olds ($M = .729, SD = .442$), independent-samples $t$ test, $t(47) = 3.491, p = .001$.

---

5 Analyzing age continuously (in months) suggests the same pattern of results. A linear regression analysis examining the effect of age in months and children’s gender (female or male) on the proportion of uninformative responses revealed a significant effect of age and gender on uninformative responses, $F(2, 71) = 14.326, p < .001, R^2 = .288$. This effect was due to a negative relation between age and children’s gender (female or male) on the proportion of uninformative responses ($t = -5.353, p < .001, B = -.024$); there was no independent effect of gender ($t = 0.314, p = .755, B = .027$).

6 We again note that analyzing age continuously (in months) suggests the same pattern of results. A linear regression analysis examining the effect of age in months and children’s gender (female or male) on the proportion of relevant responses indicated an effect of age and gender on relevant responses, $F(2, 71) = 17.783, p < .001, R^2 = .334$. This effect was due to a positive relation between age and the proportion of relevant responses ($t = 5.935, p < .001, B = .029$); there was no independent effect of gender ($t = -0.927, p = .357, B = -.087$).
Results from Experiment 1 suggest that 4- and 5-year-olds reliably form social preferences and moral judgments favoring helpers over hinderers in two scenarios previously used to evaluate infants’ sociomoral evaluations. At both ages, children preferred the helper, judged the helper to be nicer, and allocated punishment to the hinderer. When asked to justify why the selected puppet should get in trouble, 4- and 5-year-olds most often referred to the puppet’s unhelpful action during the show. In contrast, 3-year-olds did not reliably distinguish between the helper and hinderer when asked who they prefer and who was nicer. The 3-year-olds reliably identified hinderers as deserving of punishment, but approximately 50% of them did not provide informative explanations for this judgment.

Given that past research has demonstrated that preverbal infants reliably distinguish between helpers and hinderers in a forced-choice paradigm (see Hamlin, 2013b), it is somewhat surprising that 3-year-olds failed to distinguish helpers from hinderers in two of our three question types. That said, in the only other study of preschoolers’ explicit reactions to similar third-party helping and harming scenarios to date, Kenward and Dahl (2011) also observed that 3-year-olds failed to distinguish between helpers and harmers in a variety of explicit measures (but see Buon et al., 2014, for evidence that 29-month-olds can distinguish between prosocial and antisocial actors when responses to a range of explicit questions are considered together). Together, these results appear to indicate that the ability to successfully provide an implicit evaluation of helpers and hinderers during infancy is not sufficient for the provision of explicit judgments during early childhood. This pattern of performance has been observed with other types of social judgments, where infants show competence in implicit nonverbal paradigms but young children initially fail to show competence in explicit verbal ones (e.g., tasks evaluating infants’ and children’s false-belief understanding) (Baillargeon, Setoh, Sloane, Jin, & Bian, 2014; Leslie, German, & Polizzi, 2005; Wellman, Cross, & Watson, 2001).

On the other hand, it is possible that 3-year-olds’ inconsistent performance in Experiment 1 was not due to an inability to provide elicited social and moral judgments full stop but rather due to an inability to do so in our particular paradigm. For instance, 3-year-olds in Experiment 1 may have found it difficult to pay attention to the puppet shows, to process alternating events, or to remember the action performed by each puppet; indeed, Kenward and Dahl (2011) attributed 3-year-olds’ inability to distinguish between helpers and harmers when making explicit judgments to confusion regarding who did what to whom in their helping and harming scenarios. To explore this possibility, we ran an additional experiment with 3-year-olds in which children observed the very same scenarios as in Experiment 1 but with modifications designed to facilitate children’s comprehension and memory.
First, whereas in Experiment 1 helping and hindering scenarios were presented in alternation, in Experiment 2 acts were each repeated twice in a row. We reasoned that this would facilitate children’s ability to remember who did what. In addition, after viewing two acts from each puppet but before receiving test questions, children were asked comprehension questions to confirm that they remembered which puppet performed which physical action during the show. Comprehension questions required children to point to the character who opened/closed the box or gave/took the ball and so were designed only to determine whether children could identify each puppet’s action rather than to encourage children to associate these actions with positive or negative valence. If children failed to correctly respond to comprehension questions, each puppet repeated his action once more and the comprehension questions were repeated. To be included in the final sample, children needed to pass comprehension questions by the end of the third round of puppet shows/comprehension questions (maximum of four action repetitions per puppet). We predicted that these modifications would improve 3-year-olds’ performance, such that 3-year-olds in Experiment 2 would report consistent social and moral judgments favoring the helper. We also predicted that 3-year-olds in Experiment 2 would provide more informative explanations that reference the puppet’s unhelpful action and/or negative valence as the reason why he should get in trouble.

Experiment 2

Method

Participants
A total of 24 3-year-olds (M_age = 3;7, range = 3;0–3;11; 10 girls) who had not participated in Experiment 1 were tested in a university research center or the children’s day-care centers. One additional child was seen but replaced due to unwillingness to watch the puppet show.

Procedure
The only differences between Experiment 1 and Experiment 2 were that children in Experiment 2 watched two helping scenarios in a row and two hindering scenarios in a row (helping/hindering order counterbalanced) and were asked comprehension questions before test questions. Following the box scenario, children were told, “One of these guys opened the box, and one of these guys closed the box.” Children were then asked to “point to the guy that opened the box” and to “point to the guy that closed the box” (order counterbalanced). Following the ball scenario, children were told, “One of these guys gave the ball back and one of these guys took the ball away.” Children were then asked to “point to the guy that gave the ball back” and to “point to the guy that took the ball away” (order counterbalanced).

If children responded incorrectly to comprehension questions following either scenario, they were shown the helping and hindering events in that scenario again (e.g., “I don’t think he opened the box. I’m going to show you those again”) and the comprehension questions were repeated. Here, 1 child required an additional repetition of both the box and ball puppet shows, 2 children required an additional repetition of only the first puppet show (ball), and 1 child required two additional repetitions of the first puppet show (ball) before comprehension questions were answered correctly. Based on a preset criterion, 1 child’s responses following the box show were not included in the final analysis because the child responded incorrectly to comprehension questions in each of three rounds. After correctly responding to the comprehension questions, children were asked the same test questions as in Experiment 1.

Transcription and coding procedure
As in Experiment 1, two independent research assistants coded for the presence of each response type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, and nonsocial considerations). Reliability across the six categories was moderate (average Cohen's kappa = .765; see McHugh, 2012) and greatly influenced by one disagreement regarding whether a single statement reflected an irrelevant attribution. Without the irrelevant attribution category, reliability was nearly
perfect (average Cohen’s kappa = .917, range = .760–1.000). Disagreements (which involved four explanations) were resolved as in Experiment 1. Two disagreements were immediately recognized as mistakes, one regarding the irrelevant attribution disagreement described above and one due to failure to consider whether the child was responding to a box or ball show. The other two disagreements were resolved by discussion among the two coders in the presence of the first author.

Results

Social preference and moral judgments

Confirmatory analyses. As in Experiment 1, children received a score of 1 if they responded in the direction of the hypothesis for each test question (i.e., liked the helper, judged the helper as nicer, or assigned punishment to the hinderer) and 0 if they responded against the direction of the hypothesis. Children’s scores were summed across the two rounds, resulting in a score between 0 and 2 for each question type (see Fig. 2). To determine whether children responded in the predicted direction at rates exceeding chance, a series of one-sample $t$ tests were conducted comparing 3-year-olds’ liking score, niceness score, and trouble score with a chance score of 1.

In contrast to Experiment 1, 3-year-olds in Experiment 2 reliably distinguished between the helper and hinderer when providing both social and moral judgments. The 3-year-olds’ liking score ($M = 1.625, SD = 0.576$), niceness score ($M = 1.750, SD = 0.608$), and trouble score ($M = 1.667, SD = 0.637$) all exceeded chance in the direction of the hypothesis, one-sample $t$ tests comparing each score between 0 and 2 with a chance score of 1, $t(23)_{\text{liking}} = 5.318, p < .001, d = 1.085$; $t(23)_{\text{niceness}} = 6.044, p < .001, d = 1.234$; $t(23)_{\text{trouble}} = 5.127, p < .001, d = 1.047$. See Table 1 in the supplementary material for a summary of individual patterns of responding within each round.

Exploratory analyses. The differential patterns of responding among 3-year-olds in Experiment 1 and Experiment 2 suggest that the addition of comprehension questions and/or presenting the helping and hindering scenarios twice in a row rather than in alternation improved young children’s performance (it is currently unclear whether one addition played a larger role than the other; we return to this issue in the Discussion). To formally examine how judgments differed across experiments, a mixed-effects ANOVA on 3-year-olds’ responses with liking, niceness, and trouble scores as within-participants variables (repeated measures) and experiment (1 or 2) and children’s gender (female or male) as between-participants factors was conducted. This analysis revealed a main effect of experiment, $F(1, 45) = 10.095, p = .003, \eta^2_p = .183$; the 3-year-olds’ overall scores (between 0 and 6) summing across the three question types (liking, niceness, and trouble) were higher in Experiment 2 ($M = 5.042, SD = 1.488$) compared with Experiment 1 ($M = 3.600, SD = 1.683$). There were no main effects or interactions involving question type or gender (all $F$s < 1.230, all $p$s > .296).

Explanations regarding punishment

As in Experiment 1, instances of each explanation type (uninformative, relevant action, irrelevant action, relevant valence, irrelevant valence, and nonsocial considerations) were represented as proportions and averaged across the two rounds (see Table 1). We predicted that 3-year-olds in Experiment 2 would provide more informative explanations than 3-year-olds in Experiment 1. A factorial ANOVA examined the effect of experiment (1 or 2) and children’s gender (female or male) on the proportion of uninformative responses. This analysis revealed a main effect of experiment, $F(1, 45) = 4.615, p = .037, \eta^2_p = .093$, such that 3-year-olds provided fewer uninformative explanations in Experiment 2 ($M = .229, SD = .390$) compared with Experiment 1 ($M = .520, SD = .510$); there was no main effect or interaction involving gender (both $F$s < .788, both $p$s > .379).

We also predicted that 3-year-olds in Experiment 2 would provide more relevant references to the puppets’ actions or valence than in Experiment 1. To test this prediction, we combined the appeals to relevant actions and relevant valence into a single proportion of relevant responses. A factorial ANOVA

---

7 As in Experiment 1, the explanations in Experiment 2 were originally coded by two different research assistants and resulted in substantially lower reliability (average Cohen’s kappa = .733, range = .386–1.000). The coding scheme was then clarified, a more extensive training procedure was created, and the data were recoded by two different research assistants.
examined the effect of experiment (1 or 2) and children’s gender (female or male) on the proportion of relevant responses and found no difference across experiments, $F(1, 45) = 1.647$, $p = .206$, $\eta_p^2 = .035$; the 3-year-olds were equally likely to reference the puppet’s relevant action or valence when explaining their allocations of punishment in Experiment 1 ($M = .280$, $SD = .458$) and Experiment 2 ($M = .458$, $SD = .487$); there was no main effect or interaction involving gender (both $F$s < .068, both $p$s > .796). That said, references to the relevant hindering action from the puppet show were the most commonly provided type of informative explanations (53%) in Experiment 2. See Table 2 in the supplementary material for proportions of each response type following allocations of punishment to the unhelpful puppet rather than the helpful puppet.

Discussion

Overall, after correctly identifying which puppet performed which action (e.g., opened the box vs. closed the box), 3-year-olds demonstrated social preferences and moral judgments that favored the helper. Specifically, across two helping and hindering scenarios, children preferred the helper, indicated that the helper was nicer, and selected the hinderer to get in trouble. These findings suggest that 3-year-olds’ failure to reliably distinguish between the helper and hinderer in Experiment 1 was not due to an inability to interpret the helpful and unhelpful actions as socially and/or morally relevant. Rather, these results suggest that younger children in Experiment 1 had difficulty processing or remembering the helpful and unhelpful actions as presented in the first experiment, which was improved by the methodological modifications made in Experiment 2. These methodological modifications also resulted in fewer uninformative explanations when 3-year-olds in Experiment 2 justified their allocation of punishment to one puppet. To the extent that 3-year-olds provided informative explanations, they most frequently appealed to the puppet’s unhelpful action during the puppet show as the reason why the puppet should get in trouble. However, there was no difference in the proportion of appeals to relevant social considerations (actions and valence) among 3-year-olds in Experiment 1 and Experiment 2.

General discussion

Across two experiments, 3- to 5-year-olds observed third-party helping and hindering scenarios and made both social preference judgments and moral judgments. When presented with a character who helped a protagonist achieve his goal and a character who thwarted the protagonist’s goal, 4- and 5-year-olds in Experiment 1 indicated that they preferred the helper, evaluated the helper as nicer than the hinderer, and selectively allocated punishment to the hinderer. When asked to justify their allocation of punishment, children typically appealed to social considerations—primarily the selected puppet’s performance of an unhelpful action. In contrast, 3-year-olds in Experiment 1 did not provide consistent social preferences or moral judgments but instead only reliably allocated punishment to
the hinderer. Despite allocating punishment to the hinderer, 3-year-olds did not frequently refer to relevant social considerations when explaining their allocation. In Experiment 2, 3-year-olds participated in a simplified procedure where puppets’ actions were presented sequentially rather than in alternation and answered comprehension questions to assess their understanding of which puppets performed which actions. If children failed the comprehension questions, they were shown the puppet shows again. Following the successful answering of these comprehension questions, 3-year-olds also indicated that they preferred the helper, evaluated the helper as nicer than the hinderer, and selectively allocated punishment to the hinderer. When providing informative explanations regarding why the selected puppet should get in trouble, 3-year-olds most commonly appealed to social and/or moral considerations such as the puppet’s hindering action.

Although the current experiments used two of the helping and hindering puppet shows designed to explore infants’ social evaluations (see Hamlin & Wynn, 2011), there were some modifications to the presentation of the scenarios, notably the addition of a few nonvalenced words throughout the puppet show, which were added to capture preschoolers’ attention and highlight the protagonist’s goal and struggles (e.g., “Look, a toy. Too heavy!” in the box scenario), the action performed by the helper (e.g., “Open!”), and the action performed by the hinderer (e.g., “Close!”). A potential concern is that the addition of these words unduly guided children’s interpretation of the helping and hindering scenarios. We do not think this is the case for two reasons. First, the added words were not positive or negative (i.e., opening and closing are not inherently good or bad actions). Second, the added words were consistently said in a positive tone of voice to avoid suggesting that one act was more positive than another. That said, it is an empirical question how children interpret the helping and hindering scenarios without narration.

Given the similarity between the procedures previously used to examine infants’ evaluations and those currently used to examine children’s judgments, and infants’ success in forming implicit evaluations following the helping and hindering scenarios (e.g., Hamlin, 2013a, 2014b; Hamlin et al., 2011, 2013; Salvadori et al., 2015; see also Scola et al., 2015), it may be surprising that 3-year-olds in Experiment 1 did not prefer the helper over the hinderer or judge the helper to be nicer than the hinderer. However, it should be noted that there are several potentially important differences between the current procedures and those typically used with infants (beyond the addition of narration, as discussed above). Specifically, infants are exposed to more instances of the helping and hindering scenarios (a minimum of six total events as opposed to the four total events in Experiment 1). Infants also observe the scenarios from the end of a long table with a curtain blocking the experimenter from view. In the current experiments, the scenarios were enacted directly in front of children with the experimenter fully visible. Thus, it is possible that 3-year-olds in Experiment 1 would have provided consistent social and moral judgments if they observed more instances of the scenarios and/or were tested without the potential distractions of the nearby puppets and/or experimenter.

Further modifications to the helping and hindering puppet show paradigm were introduced in Experiment 2. These modifications significantly improved 3-year-olds’ performance, resulting in consistent social and moral judgments and more interpretable explanations regarding their allocation of punishment (cf. Kenward & Dahl, 2011). It is currently unclear which aspect(s) of these modifications is responsible for this improvement: the sequential rather than alternating presentation of helping and hindering scenarios, the answering of comprehension questions regarding each puppet’s action, or the repetition of the helping and hindering scenarios after incorrect responses to comprehension questions. Although not definitive, it seems unlikely that 3-year-olds’ performance improved due to increased exposure to the helping and hindering scenarios. Very few children answered comprehension questions incorrectly on the first round; thus, on average, children saw 4.26 acts in Experiment 2 versus 4 acts (2 helping and 2 hindering) in Experiment 1. In addition, we also feel that it is unlikely that 3-year-olds’ improved performance in Experiment 2 is due to answering comprehension questions. In another study recently completed in our lab where actions were presented sequentially rather than in alternation but comprehension questions were not used, 3-year-olds acted like 3-year-olds in the current Experiment 2, providing reliable moral judgments regarding helpers and hinderers (Van de Vondervoort & Hamlin, 2017). Overall, then, it seems that the sequential rather than alternating presentation of helping and hindering scenarios in Experiment 2 is likely responsible for the difference in 3-year-olds’ responding across Experiments 1 and 2, presumably because sequential
presentation facilitates processing of which puppet performed which action. That said, more work is necessary to determine how young children process and interpret these helping and hindering scenarios.

These studies contribute to a growing literature regarding children's responses to morally valenced actions. Previous work has demonstrated infants' preferences for prosocial rather than antisocial others during the first year of life (e.g., Hamlin & Wynn, 2011; Hamlin et al., 2007). By 3 or 4 years of age, children clearly distinguish between moral transgressions and conventional or personal transgressions (see Smetana, 2006, for a review) and spontaneously protest the performance of antisocial behaviors (e.g., Rossano et al., 2011). The current experiments suggest that 3- to 5-year-olds evaluate goal facilitation and goal blocking as morally relevant behaviors and form a variety of moral judgments (i.e., positive moral evaluations and allocation of punishment) based on these actions. Furthermore, when asked to justify a moral judgment (the allocation of punishment), children often appeal to the performance of unhelpful actions within the hindering scenario. Although it is possible that appeals to the hindering action from the puppet show reflect nonsocial concerns (e.g., reports regarding the physical action performed) or social but not moral concerns (e.g., reports regarding the puppet's role in blocking the protagonist's goal without any sense that this action was inappropriate), it is also possible that these explanations reflect moral concerns. Specifically, appeals to the puppet's hindering action when explaining why the puppet should get in trouble may suggest that children's moral judgments have a normative basis—that the puppet ought not have blocked the protagonist's goal and so should be punished—rather than a nonsocial or idiosyncratic social basis.

In the current experiments, we asked children to only explain their moral judgments to prevent explanations regarding moral judgments from influencing explanations regarding social preferences and vice versa. To further explore the cognitions underlying preferences for helpful versus unhelpful others, future work should explore children's justifications for their social preferences. To illustrate, it would be informative to determine the extent to which children appeal to nonsocial considerations when explaining social preferences such as idiosyncratic preferences based on the puppet's appearance or neutral aspects of the puppet show display (e.g., nonhelping or hindering actions performed during the show). If children are more likely to appeal to social considerations when explaining moral judgments and to appeal to nonsocial considerations when explaining social preferences, this would suggest that social and moral responses to the helping and hindering scenarios are (or can be) distinct.

In addition, although the methods used here were chosen based on a host of previous work in the moral development literature (e.g., Baird & Astington, 2004; Costanzo et al., 1973; Cushman et al., 2013; Leslie et al., 2006; Zelazo et al., 1996), it would also be fruitful to explore children's reactions to helping and hindering using other common moral development methodologies. For example, our methods could be modified to examine children's spontaneous protests to helpful versus unhelpful acts (e.g., if puppets were to warn about what they were going to do before they do it, children might step in and instruct them not to do so) (e.g., Rossano et al., 2011; Vaish et al., 2011) or to examine children's assessments of whether it was “okay” versus “not okay” for a puppet to behave as he did given the existence of authority sanctions, unique contexts, and the like (see Smetana, 2006, for a review). Of note, a judgment regarding whether a single puppet's actions are okay or not okay is distinct from the forced-choice judgments that preschoolers provided in the current experiments. Although categorical judgments that the helper's actions were acceptable and the hinderer's actions were unacceptable would be consistent with the findings of the current experiments, it is an open question whether children provide categorical moral judgments following helping and hindering scenarios. These methods would further help to elucidate the social versus moral nature of children's reactions to helpers and hinderers.

A final open question is whether infants' implicit preferences for helpers over hinderers are related to children's explicit social and moral judgments. A fruitful area of future research would be to explore the continuity of moral evaluation and behavior across the moral development of individual children, beginning with puppet preferences during infancy. Are those children who were more likely to choose prosocial puppets during infancy also more likely to identify prosocial puppets as nicer and antisocial puppets as deserving of punishment during childhood? Do they perform better on other measures of social and/or moral functioning? Establishing whether there is within-individual continuity from infancy through childhood would greatly add to the debate regarding whether infants' responses
should be considered precursors to a moral sense. In sum, the current experiments shed light on preschoolers’ explicit evaluations regarding third-party helpers and hinderers and suggest multiple avenues for future research.

Acknowledgment

This work was supported by a grant from the Social Sciences and Humanities Research Council (SSHRC 435-2014-2173) awarded to J. Kiley Hamlin.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2017.07.004.

References


