Developmental Psychology

Infants' Evaluation of Prosocial and Antisocial Agents: A Meta-Analysis

--Manuscript Draft--

Manuscript Number:	DEV-2017-0934R2
Full Title:	Infants' Evaluation of Prosocial and Antisocial Agents: A Meta-Analysis
Abstract:	In the last decade, numerous studies reported that infants prefer prosocial agents, who provide help, comfort or fairness in distributive actions, to antisocial agents, who harm others. We meta-analyzed the results of published and unpublished studies on infants aged 4-32 months and estimated that approximately two infants out of three, when given a choice between a prosocial and an antisocial agent, choose the former. This preference does not change with age and is consistent across other factors, such as the type of dependent variable (selective reaching or helping) and the modality of stimulus presentation (cartoons or real events). Effect size is affected by the type of familiarization events: giving/taking actions increase its magnitude compared to helping/hindering actions. There is evidence of a publication bias, suggesting that the effect size in published studies is likely to be inflated. Also, the distribution of children who chose the prosocial agent in experiments with N = 16 suggests a file drawer problem.
Article Type:	Article
Keywords:	meta-analysis; infancy; social cognition; helping; moral goodness; fairness
Corresponding Author:	Francesco Margoni University of Trento, Italy ITALY
Corresponding Author E-Mail:	francesco.margoni@unitn.it
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Trento, Italy
Other Authors:	Luca Surian
Corresponding Author's Secondary Institution:	
First Author:	Francesco Margoni
Order of Authors Secondary Information:	
Suggested Reviewers:	Kiley Hamlin University of British Columbia kiley.hamlin@psych.ubc.ca A substantial part of the studies we meta-analyzed were published by Kiley Hamlin and her colleagues.
	Paul Bloom Yale University paul.bloom@yale.edu He is one of the leading expert on infants moral sense.
	Kristen Dunfield Concordia University kristen.dunfield@concordia.ca She has done important work on helping and prosociality in infancy.
	Marine Buon University College London marinebuon@gmail.com She has done relevant work on moral reasoning both with infants and adults.
Opposed Reviewers:	
Order of Authors:	Francesco Margoni
	Luca Surian

Infants' Evaluation of Prosocial and Antisocial Agents: A Meta-Analysis

Francesco Margoni and Luca Surian

Department of Psychology and Cognitive Sciences, University of Trento

Authors note:

Francesco Margoni, Department of Psychology and Cognitive Sciences, University of Trento; Luca Surian, Department of Psychology and Cognitive Sciences, University of Trento.

Correspondence should be addressed to Francesco Margoni, Department of Psychology and Cognitive Sciences, University of Trento, Corso Bettini 31, Rovereto (TN), 38068, Italy. E-mail: francesco.margoni@unitn.it

Infants' Evaluation of Prosocial and Antisocial Agents: A Meta-Analysis

Second resubmission: 12 October 2017

Abstract

In the last decade, numerous studies reported that infants prefer prosocial agents, who provide help, comfort or fairness in distributive actions, to antisocial agents, who harm others. We meta-analyzed the results of published and unpublished studies on infants aged 4–32 months and estimated that approximately two infants out of three, when given a choice between a prosocial and an antisocial agent, choose the former. This preference does not change with age and is consistent across other factors, such as the type of dependent variable (selective reaching or helping) and the modality of stimulus presentation (cartoons or real events). Effect size is affected by the type of familiarization events: giving/taking actions increase its magnitude compared to helping/hindering actions. There is evidence of a publication bias, suggesting that the effect size in published studies is likely to be inflated. Also, the distribution of children who chose the prosocial agent in experiments with $N = 16$ suggests a file drawer problem.

Keywords: meta-analysis, infancy, social cognition, helping, moral goodness, fairness

Introduction

Adults routinely praise prosocial actions and blame antisocial actions. The first studies on the early-emerging abilities to generate and express socio-moral evaluations presented preverbal infants with a scenario in which someone helped and someone else hindered a target protagonist (Hamlin, Wynn, & Bloom, 2007; see also Kuhlmeier, Wynn, & Bloom, 2003). In Hamlin et al. (2007), infant saw a character (a wooden circle, with large googly eyes) that repeatedly attempted to climb a hill. A prosocial character (e.g., a triangle) and an antisocial character (e.g., a square) were then introduced. The former helped the climber to climb the hill, but the latter pushed the climber down. When asked to pick up one of these two wooden blocks with plastic large googly eyes, infants reached for the prosocial agent and picked the helper rather than the hinderer.

This initial result attracted the attention of the scientific community, and a host of studies were conducted to investigate infants' socio-moral preferences and expectations about situations involving morally good and bad agents (Bloom & Wynn, 2016; Van de Vondervoort & Hamlin, 2016; for a review see Holvoet, Scola, Arciszewski, & Picard, 2016). This line of research aimed at unveiling the origins of our moral understanding and capacity to generate and express moral evaluations. In this body of research, infants' evaluations are inferred from infants' preferences for prosocial over antisocial agents. However, it is an open question whether these preferences reveal evaluations that are, to an interesting degree, isomorphic to the moral evaluations generated by older children or adults (who are also able to motivate and justify their judgments), or whether they should be tentatively considered as an initial building block for the later development of prescriptive reasoning and evaluative rules (Killen & Smetana, 2015).

To date, most studies presented infants with successful acts of helping and hindering (e.g., Hamlin et al., 2007; Salvadori et al., 2015), but other studies presented a failed attempt

to help or hinder (e.g., Hamlin, 2013; Lee, Yun, Kim, & Song, 2015) or events that focused on fairness in distributive actions, another central part of the moral domain (e.g., Burns & Sommerville, 2014; Geraci & Surian, 2011; Sloane, Baillargeon, & Premack, 2012).

Two main kinds of tasks were employed in the study of infants' socio-moral cognition. First, there are *attentional tasks*, in which the child is required to watch a sequence of events, and his or her looking times at test events are recorded. These tasks follow the 'violation of expectation' paradigm and rely on the assumption that infants tend to look longer at events violating, rather than confirming, their expectations. By employing this measure, researchers demonstrated that, by the second year of life, infants expect others to distribute resources equally if recipients are equally deserving (Meristo, Strid, & Surian, 2016; Schmidt & Sommerville, 2011; Ziv & Sommerville, 2016), or according to merit if recipients are differentially deserving (Buyukozer Dawkins, Sloane, & Baillargeon, 2016; Surian & Franchin, 2017a; Sloane, Baillargeon, & Premack, 2012; see also Baillargeon, Scott, He, Sloane et al., 2015 for a review). Attentional tasks may also yield two additional dependent variables: (a) *anticipatory looks*, which help to understand whether the child anticipated a certain action outcome by recording whether she first looked in a certain area of interest (e.g., Geraci & Surian, 2011) and (b) *preferential looking times*, which allow to assess whether children prefer to look at one agent over another agent (e.g., Hamlin, Wynn, & Bloom, 2010).

The second set of tasks used in the study of infants' and toddlers' socio-moral evaluations, the *manual tasks*, require the child to act by manually choosing, pointing, helping, or giving a treat. Overall, the studies that used these paradigms reported evidence of an early-emerging preference for the helper over the hinderer (Hamlin, 2015a; Hamlin & Wynn, 2011; Hamlin et al., 2007), for someone who showed a helping intention over someone who showed a hindering intention (Hamlin, 2013; Hamlin, Ullman, Tenenbaum,

Goodman et al., 2013), for a fair distributor over an unfair distributor (Burns & Sommerville, 2014; Geraci & Surian, 2011), for a giving agent over a taking or keeping agent (Hamlin, 2014; Hamlin, Wynn, Bloom, & Mahajan, 2011; Scola, Holvoet, Arciszewski, & Picard, 2015), and for a comforting agent over an harming agent (Buon et al., 2014; see also Kanakogi, Okumura, Inoue, Kitazaki et al., 2013). The current meta-analysis focuses on this last group of studies, which investigated infants' socio-moral preferences with manual tasks. Here we were interested in estimating infants' evaluations of prosocial and antisocial agents, as these are expressed by a preference. Therefore, we did not take into account studies that measured only infants' looking times, which, in most cases, investigated children's expectations, not their preferences. Moreover, taking into account also the few studies that used looking times to assess infants' preferences would have resulted in increasing the sample heterogeneity of the selected studies.

The present study

The aim of the present study was threefold. First, we wished to provide an estimate of infants' preference for prosocial agents by analyzing all the effect sizes found in published and unpublished studies. Second, we aimed to assess whether infants' age has an influence on the estimated effect sizes, and so whether developmental changes in the ability to express socio-moral preferences occur. Third, we addressed a number of ancillary questions such as whether the sample size, the type of scenario, the modality of stimulus presentation, the type of dependent variable, and the agents from which children were encouraged to choose exert an influence over the effect size estimate, or whether there is evidence of a publication bias.

With respect to the first aim, although a wealth of studies reported a significant preference for helpful or fair agents over hindering or unfair agents, some recently published and unpublished works failed to replicate these findings (e.g., Abramson, Dar, Te'eni, & Knafo, 2016; Salvadori, Blazsekova, Volein, Karap et al., 2015). Thus, a meta-analysis of the

existing evidence seems useful to cast light of this phenomenon. To start, Scarf, Imuta, Colombo, and Hayne (2012) noted that in Hamlin et al. (2007)'s study the climber bounced after reaching the top of the hill, but he did not bounce after being hindered and pushed at the bottom of the hill. They suggested that infants' choices were driven by the bouncing event, rather than by the moral status of the agents. In fact, in their study, infants chose the helper when the climber bounced at the top of the hill, but chose the hinderer when the climber bounced at the bottom of the hill. However, a subsequent study clarified that in Scarf et al. (2012) infants' choices were driven by the bouncing event because the characters were presented without adequate physical and behavioral cues (Hamlin, 2015a). If, unlike in Scarf et al. (2012), (a) the climber has eyes with fixed pupils that look uphill, and therefore the agent does not look down for the entire event; (b) the speed of the climber's movements varies, and therefore ascending and descending movements do not seem equally intentional; and (c) at the end of the event, the climber does not move up or down the hill on his own, but he is clearly either pushed by the helper or by the hinderer, infants' preferences are again driven by the socio-moral nature of the interactions rather than by perceptual cues. This finding suggested that the failure to replicate reported by Scarf et al. (2012) was due to small but crucial differences in the experimental stimuli.

However, five other studies failed to find a preference for the prosocial agent in early infancy. Salvadori and colleagues (2015) employed with 9-month-olds the same stimuli and procedure used by Hamlin & Wynn (2011), who reported a preference for the helping puppet, but did not find any preference. Another replication failure was reported by Cowell & Decety (2015), though methodological differences with previous studies should be taken into account, in particular the length of the testing sessions and the procedures required to collect EEG data, that may have been caused distraction or fatigue in infants. Moreover, three other unpublished studies found no evidence of a preference for helping over hindering agents (at 7

and 12 months: Hamlin, 2016; at 8 and 10 months: Woo & Hamlin, 2016; and at 9 and 18 months: Abramson et al., 2016).

Given these findings, the question of what is the average effect size of infants' preference for prosocial over antisocial agents remains currently unanswered, and we may wonder whether unpublished studies significantly influence the estimation of the effect size. In fact, we could predict that published studies report larger effect sizes than unpublished studies, as scholars often find difficulties in publishing negative results and replications, or these are in any case under-reported (Borenstein, Hedges, Higgins, & Rothstein, 2011; Laws, 2013). Meta-analysis is a well-known and powerful statistical tool that returns a reliable and trustworthy synthesis of the current evidence, by providing an estimate of both the dimension and the precision of the effects we chose to consider (Crocetti, 2016; Cumming, 2012, 2014).

Meta-analyses are useful also because they allow scholars to assess whether some factors explain the between-studies variability in the effect sizes. Thus, they favor hypothesis testing on a large amount of data. One crucial issue concerning socio-moral preferences during early infancy is whether age significantly affects the likelihood to choose the prosocial agent. In the current meta-analysis, we included experiments conducted on infants between 4.5 and 32 months of age. Four months is the youngest age at which a preference can be manifested by a reaching behavior (Lobo & Galloway, 2013; McDonnell, 1975; Thelen, Corbetta, Kamm, Spencer et al., 1993; von Hofsten, 1984). This meta-analysis is thus useful in assessing whether there is an early-emerging preference for prosocial agents, and whether developmental changes in this preference can be observed in infants aged 4 to 32 months. Finding no age effects would indicate that the social input infants are exposed to and the general changes occurring between 4 and 32 months are unrelated to infants' ability to express socio-moral preferences in the tasks used in the selected literature.

Furthermore, we assessed whether some other possible predictors influence the

estimated average effect size. First, we investigated whether the sample size of the studies affects the estimated effect size. It is widely understood that small sample sizes risk undermining the accuracy of the models. As sample size increases, the precision of estimates also increases until reaching an asymptote when the maximum accuracy is achieved (unless one has the possibility to directly describe the population by studying each individual). Studies with small sample sizes are known to be susceptible to inflated effect size estimates and, if constituting the only evidence available, they are likely to bias the estimation of the true effect size (Button et al., 2013; Oakes, 2017; see also Ioannidis, 2005). Therefore, testing the influence of sample size on effect size is useful in assessing the accuracy of the estimation.

Second, we tested the generalizability of the results across different domains or examples of morality, such as fairness, helping or hindering, giving or taking. We assessed whether different types of actions (e.g., fair distributions as opposed to giving or helping actions) differentially affect children's choices. To further investigate the generalizability of the findings, we assessed the impact of the type of the dependent variable (whether the child was asked to reach for one character or to help a character) and the type of agent the child was presented with during the manual task (it could be puppets, foam shapes or two experimenters) on the likelihood of choosing the prosocial agent. Investigating whether these factors have an impact on the effect size estimation would help to assess the stability and pervasiveness of the phenomenon.

Finally, we asked whether presenting infants with cartoons vs. real events (both live shows and movies displaying real interactions between experimenters or puppets) has a significant impact on the child's understanding of the presented events and influences the socio-moral choice. On the one hand, our phylogenetic history has presented us with real individuals, and only recently with 2D graphics. Therefore, our mind, especially during

infancy, may be wired to real exchanges between people, and the child's processing and understanding could be enhanced by real events rather than cartoons. On the other hand, cartoons may be easier to process because they typically present stimuli that are perceptually simpler than real events or movies displaying real interactions. Also, we asked whether infants' preferences are affected by whether children are presented with live shows as opposed to movies. Perhaps, live shows enhance children's understanding by making the interactions between agents more interesting to watch.

Method

Literature search and inclusion criteria

We initially searched for relevant studies by consulting the commonly used electronic database PsycINFO. We conducted a full-text search using the terms *infant**, *moral**, *help**, *hinder**, *good**, or *fair** (the search was conducted in November, 2016[1]). We selected articles on the basis of the following three criteria: a) the study assessed infants' choices with a manual task, either using a manual choice paradigm, a selective helping paradigm or a paradigm in which the child is encouraged to offer some goods to one of the two characters—we therefore excluded studies that used attentional tasks and measured only looking times or preferential looking[2]; b) infants were given the opportunity to express their preference between a 'morally good' and a 'morally bad' character, where a good character can be a helping agent, a fair distributor, or a giver agent, and a bad character can be a hindering, unfair, or keeper/taker agent, respectively; c) the study was conducted on infants and toddlers between 4 and 36 months of age.

After this initial research, we followed up by looking for any relevant study in the

[1] In June 2017, in reviewing the manuscript for publication, we updated two references that in the original submission were unpublished (Steckler et al., 2017; Surian & Franchin, 2017b).
[2] One study employed only preferential looking to assess (3-month-old) infants' preferences (see Hamlin et al., 2011). In order to maintain the sample homogeneity, we did not include this study in the current meta-analysis.

references lists of the retrieved articles. We then requested any unpublished data by posting a call on a few online forums, and by consulting experts in the field. Moreover, we searched for unpublished works within the programs of a few main conferences on infant cognition, after 2007, that is, the year of publication of the seminal work by Hamlin and colleagues. At the end of this literature search, we collected 26 studies, containing 61 relevant effect sizes (44 published, 17 unpublished), with a total *N* of 1244 participants (see Table S1 in Supplemental Materials for a detailed list of the studies, and Figure 1).

Coding and effect sizes calculation

Two raters that were not blind to the aims of the meta-analysis (the authors) read the studies and independently coded their methodological characteristics and outcomes. Each rater decided whether the experiments fulfilled the inclusion criteria. Overall, inter-rater agreement initially averaged 86% for all coded factors, and all the disagreements that arose were then resolved by discussion.

We coded the sample mean age and size. Since Hamlin and her colleagues collected half of the data (30/61 effect sizes), and their work has recently been followed up by some failed attempts to replicate (e.g., Salvadori et al., 2015), we also coded whether the study was conducted in their laboratories. Finally, we coded the type of scenario, the modality of stimulus presentation, the dependent variable (whether infants were asked to pick up an agent or selectively help her), and the number of infants that chose the prosocial agent over the antisocial agent. With respect to the type of scenario, we coded: 1) whether infants saw helping and hindering events (both simple and complex cases, such as failed attempts to help or hinder), or fair and unfair actions, or events of giving or returning an object to others vs. taking an object from others, or other actions (e.g., harm vs. comfort); and 2) whether they were presented with a real event (live shows and movies staging puppets or human actors) or cartoons—we also coded whether infants were presented with a live show or movies (the

wooden version of the hill by Hamlin et al. was coded as a live show). With respect to the dependent variable, we specified whether infants were invited to reach for an agent (or a toy offered by the agent) or to help her (giving a treat and share were considered acts of helping or pleasing; only one experiment in Hamlin et al. 2011 encouraged infants to give a treat to one of the puppets, and only Enright and Sommerville 2016 asked infants to share with one of the puppets). We also coded whether infants were asked to show a preference between experimenters (or toys offered by experimenters, in two studies), puppets or foam shapes.

We conducted the analyses with the R package *Metafor* (Viechtbauer, 2010). Before conducting the analyses, using the function *escalc*, we converted all the results in a couple of values: a logit transformed proportion (of children that chose the prosocial agent), and the associated sampling variance. We used the logit transformed proportion instead of a raw proportion to respect the normality assumption of the meta-analysis. However, for ease of interpreting, in the Figures we reported effect sizes calculated as raw proportions (number of children that chose the good agent divided by sample size), without the logit transformation.

Results

Our main aim was to calculate an average point estimate for the size of infants' preference towards prosocial agents, as well as its 95% confidence interval. Twenty-six published and unpublished reports yielded 61 effect sizes (see Table S1). We first conducted a fixed-effects meta-analysis, which assumes that all studies share a common true effect, and does not take into account between-study variability. The estimate of the common effect size was 0.62, 95% CI (0.50, 0.75). We then ran a random-effects meta-analysis, with $k = 61$ and a restricted maximum-likelihood estimation for τ^2 (estimated amount of total heterogeneity between studies, that is, the between study variance). Here, a random-effects approach is more appropriate than a fixed-effects approach because in a random-effects model both within- and between-study variability are taken into account. While a fixed-effects model

describes *k* studies, a random-effects model takes the *k* studies as a sample from a larger population (Card, 2009). The estimated average logit transformed proportion was 0.75, 95% CI (0.56, 0.94); when, for ease of interpreting, it is transformed backwards to raw proportion, it is equal to 0.68, 95% CI (0.64, 0.72)[3]. Thus, about two out of three children chose the prosocial over the antisocial agent.

The estimated amount of total heterogeneity (τ^2) was equal to 0.27, 95% CI (0.13, 0.63), and the Higgins' I^2 (which estimates the percentage of how much of the total variability in the effect size estimate can be attributed to the heterogeneity among the true effects) was 52%, CI (35%, 71%), which suggests moderate heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). We also found a significant heterogeneity between studies, $Q(60) = 127.56$, $p < 0.001$, and this motivated us to further assess the influence of potential moderators.

[3] The calculation performed directly on the raw proportions yielded similar results.

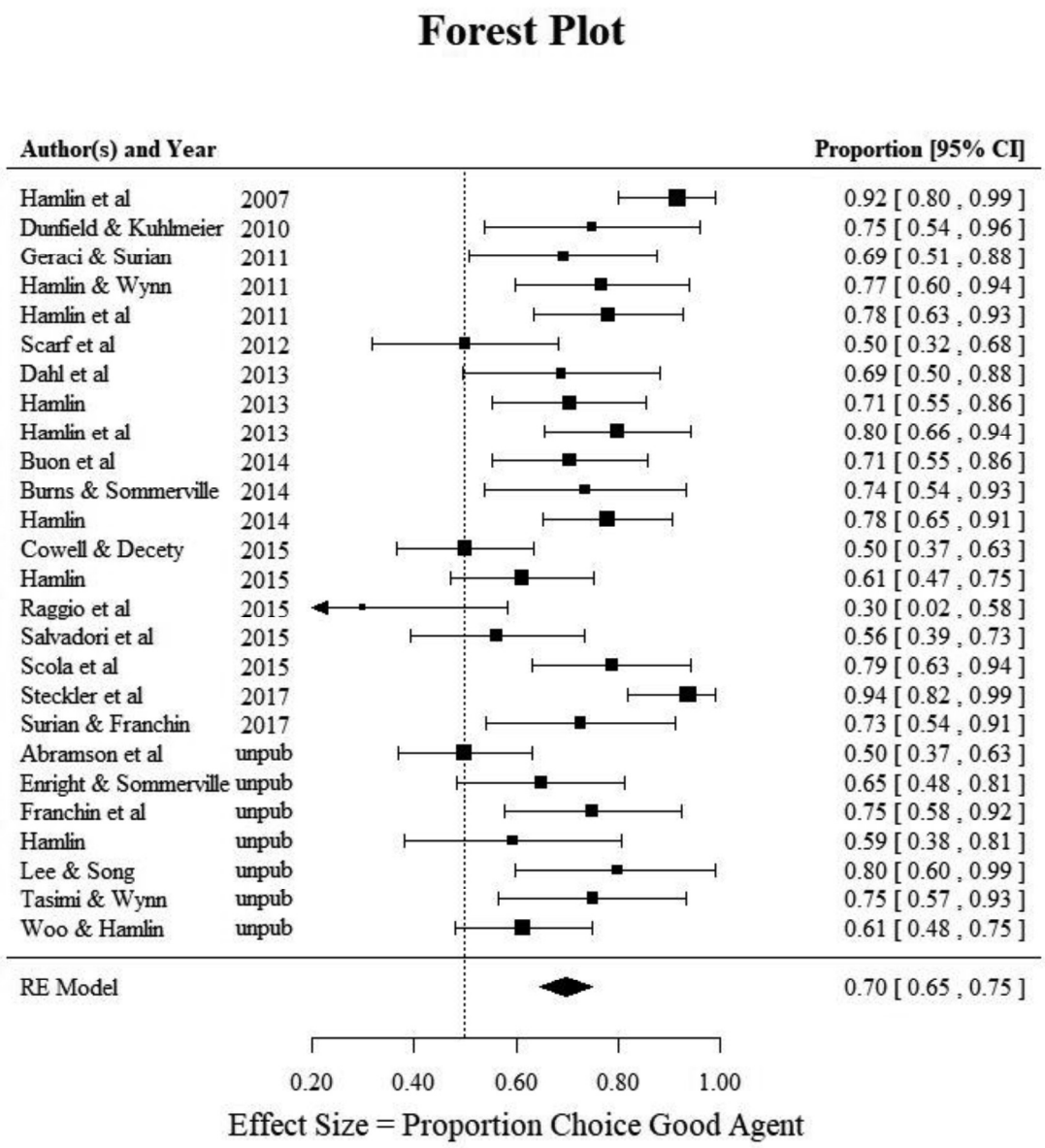

Figure 1. Forest plot showing the aggregate raw proportions and the 95% confidence intervals for all the studies included in the meta-analysis. For ease of reading, here we show the estimated average effect size computed from the displayed dataset of $k = 26$, where each row lists the aggregate effect size from all the relevant effects found in the cited study. The vertical dotted line at 0.50 represents the reference point indicating no preference or random response.

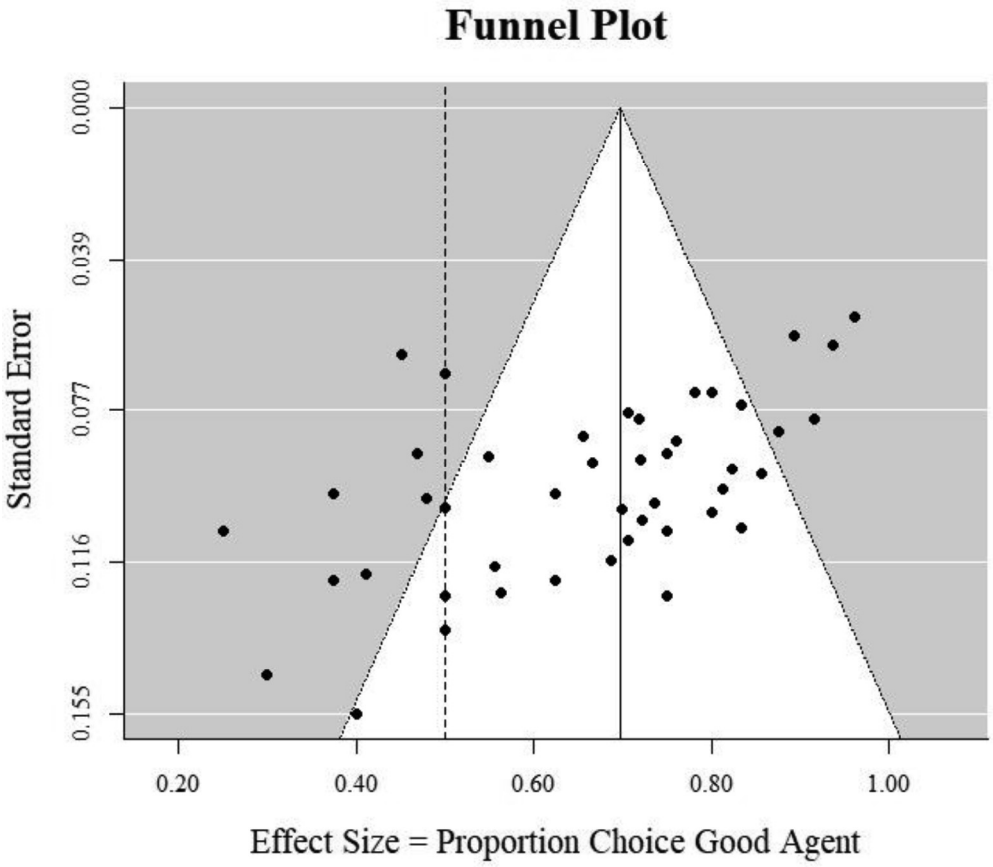

Figure 2. Funnel plot showing standard errors as a function of effect sizes. For ease of interpreting, we displayed the raw proportions. Chance level (0.50) is represented with a dashed vertical line. The outer dotted lines delimit the triangular region within which 95% of effect sizes are expected to lie when biases and heterogeneity are absent. The graph shows a percentage of effect sizes falling outside the triangular region greater than 5%, suggesting a moderate heterogeneity.

Assessing the publication and 'laboratory bias'

Part of the heterogeneity among studies may be due to the influence of moderators. A first predictor could be whether the study was published or not. In fact, the estimated average raw proportion of infants that chose the prosocial agent in the retrieved unpublished studies is

64%, while the proportion in published studies is 70%. We thus conducted a mixed-effects meta-analysis, which allows us to explain the variability by taking into account also some study-level factors. We ran the model with $k = 61$ and a maximum-likelihood estimation for τ^2 (as suggested when comparing two or more fixed-effects models; see Faraway, 2006, Chapter 8). However, while the fact that the study was published corresponded to an increase of 5.64 units of average raw percentage of choosing the prosocial agent, 95% CI (4.66, 6.58), the analysis did not show a significant impact of the moderator (whether the effect size was published or not), $QM(1) = 1.66$, $p = 0.197$.

We then asked whether there is a publication bias in the literature on infants' socio-moral preferences (Ferguson & Brannick, 2012). To answer this question, we modified the dataset and included only the effect sizes of published studies, $k = 44$. Using the trim and fill method (Duval & Tweedie, 2000), we estimated 9 missing studies on the left side of the effect size distribution to produce the expected symmetric funnel plot. Moreover, the calculation that accounted for the publication bias and that was performed on the dataset with only published studies returned that the estimated average logit transformed proportion adjusted for the publication bias was 0.65, 95% CI (0.40, 0.90), that backwards transformed in raw proportion is 0.66, 95% CI (0.60, 0.71). An Egger's test for asymmetry (using the standard error as predictor; Egger, Smith, Schneider, & Minder, 1997) confirmed the presence of a publication bias, $Z = 4.39$, $p < 0.001$.

In sum, the trim and fill procedure and the Egger's test showed evidence of a publication bias that causes in published studies an overestimation of the proportion of children picking the prosocial agent. However, we also reported that published studies do not report significantly bigger effect sizes than the set of unpublished studies we managed to retrieve. Therefore, it is certainly possible that the current meta-analysis does not include all the existing unpublished studies. In fact, using the trim and fill method on the entire dataset (k

= 61), we again estimated 11 missing studies on the left side of the effect size distribution to produce the expected symmetric funnel plot, and the calculation accounting for the publication bias estimated an average logit transformed proportion of 0.59, 95% CI (0.40, 0.78), that backwards transformed in raw proportion is 0.64, 95% CI (0.60, 0.69). The Egger's test confirmed that a publication bias remains also when including the unpublished studies, $Z = 5.68$, $p < 0.001$.

Another question we addressed, in light of the recent failed attempts to replicate Hamlin and colleagues' findings (e.g., Salvadori et al., 2015), was whether the effect sizes reported by Hamlin et al. are similar to the effect sizes reported by other researchers. The mixed-effects model with the Hamlin's Lab moderator (yes or no) showed a tendency toward significance, QM(1) = 3.24, $p = 0.072$. The average raw proportion of infants preferring the prosocial agent was 5.83 points higher than in studies conducted by other laboratories, 95% CI (4.93, 6.69). Hamlin and her collaborators reported, on average, a proportion equal to 0.71, whereas the other researchers reported an average proportion of 0.64. To further address the point, since Hamlin and colleagues did not investigate infants' preferences for fair over unfair agents, we excluded from the dataset the studies on fairness and ran a mixed-effects model with $k = 47$ (only help/hinder and give/take effect sizes), and 'Hamlin's Lab' as a moderator. The analysis showed a significant impact of the moderator, QM(1) = 5.22, $p = 0.022$. Being reported by Hamlin et al. corresponds to an increment of the effect size equal to 6.29 units in terms of the average raw proportion of preferring the prosocial agent, 95% CI (5.19, 7.27).

Age and sample size as potential moderators

Sample mean ages ranged from 139 days (4.6 months) to 960 days (32 months; $M = 390$, $SD = 213$). The effect size of infants' preferences did not depend on their age (see Figure 3a). By running a mixed-effects model with $k = 61$, maximum-likelihood estimation

for τ^2, and age as a moderator, we did not find a significant influence of age on effect size, QM(1) = 0.02, p = 0.892. Moreover, age did not explain any proportion of the observed heterogeneity in the effect sizes, R^2 < 0.01%.

By contrast, sample size accounted for a significant proportion of the observed heterogeneity, R^2 = 21.26%, Q(1) = 4.61, p = 0.032. Experiments with larger samples reported smaller effect sizes, and vice versa; an increase of one unit in the sample size corresponds to a decrease of 0.49 units in terms of the average raw proportion of choosing the prosocial agent, 95% CI (0.49, 0.50). However, this effect was due to two outliers: Cowell & Decety (2015) that had a N = 54 (z-score = 3.62), and Abramson et al. (2016, Sample B) that had a N = 62 (z-score = 4.48). After excluding these two items, that had a N three and four standard deviations above the mean, sample size did not account for a significant proportion of heterogeneity, R^2 = 0.33%, Q(1) = 0.18, p = 0.671 (see Figure 3b).

Note that 24 effect sizes out of 61 came from conditions that had N = 16, and, among them, eight (so, one-third) reported that 12 infants chose the prosocial agent, that is, the minimum number of successes needed to reach statistical significance in a one-tailed test. We asked whether the distribution of successes of these 24 experiments with N = 16 differed from the theoretical binomial distribution. A Binomial dispersion test showed that the observed distribution was significantly different from the expected distribution computed by setting *p(success)* = 0.71 and N = 16, $\chi^2(1, N = 24) = 50.37$, $p < 0.001$.

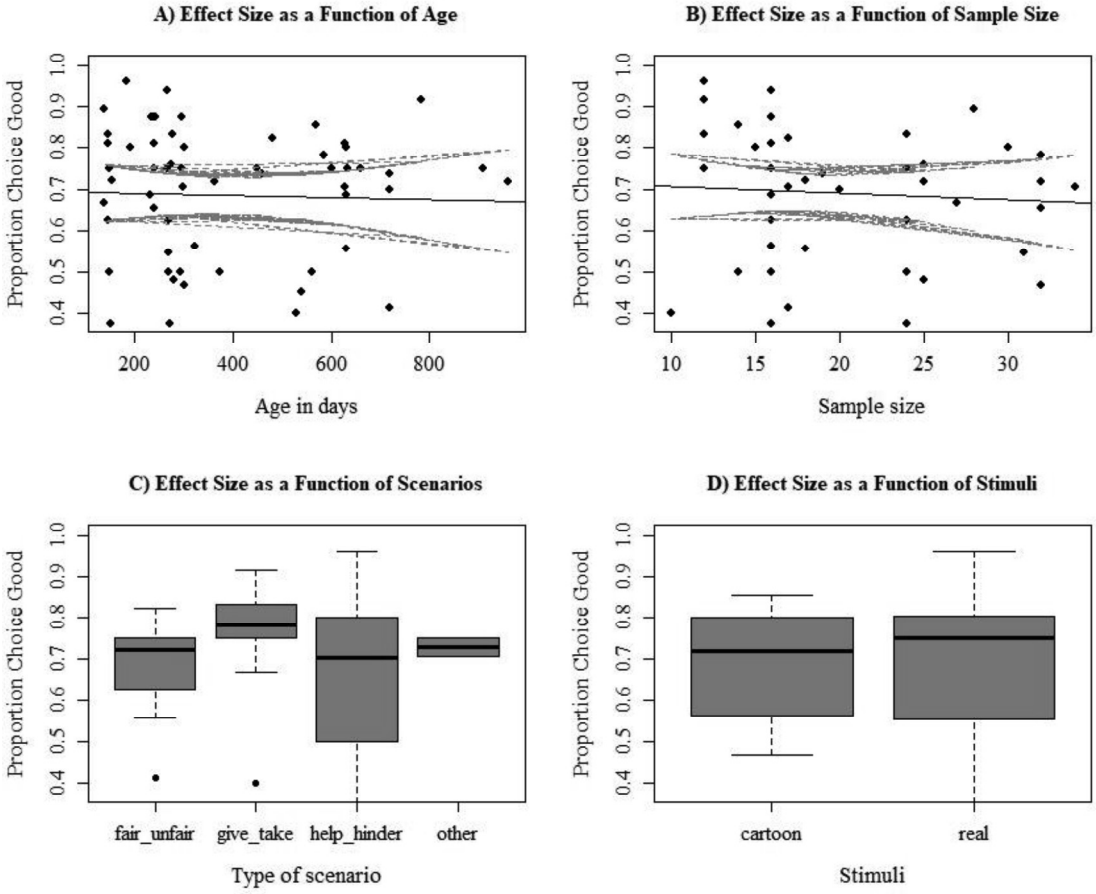

Figure 3. Effect sizes as a function of age in days (a), sample size (b), type of scenario (c), and modality of stimulus presentation (d). For ease of interpreting, we displayed raw proportions. In graphs (a) and (b), the grey lines represent the confidence intervals at 95% for each effect size. In graph (b), the two outliers (Cowell & Decety, 2015; Abramson et al., 2016, Sample B) were not displayed.

Other moderators

Do the type of scenario, the modality of stimulus presentation, the type of dependent variable, and the type of agents from which infants were asked to choose have an influence over their preferences? In order to answer this question, we inserted these categorical variables as moderators in four separate mixed-effects models, with $k = 61$ and maximum-

likelihood estimation for τ^2. The Scenarios variable had four levels (help/hinder, fair/unfair, give/take, other), Stimuli had two levels (real vs. cartoon, or, in a second analysis, live show vs. movies), Measure had two levels (reach, help), and Agents had three levels (experimenters, puppets, simple geometrical shapes).

Scenarios had a significant influence on infants' preferences, QM(3) = 7.86, p = 0.049, R^2 = 30.82% (see Figure 3c). Then, since 34 effect sizes fell into the help/hinder category, 12 into the fair/unfair, 13 into the give/take, but only two fell into the 'other scenario' category, we performed again the analysis without the latter category, and again found that Scenarios influenced infants' preferences, QM(2) = 7.17, p = 0.028, R^2 = 28.21%. The estimated average raw proportion of infants choosing the prosocial agent after having observed give/take events is 77%, after fair/unfair events is 69%, and after help/hinder events is 63%. In particular, the estimated proportion associated with give/take scenario was significantly higher than the estimated proportion associated with help/hinder scenario, QM(1) = 7.44, p = 0.006. Next, using the same dataset (k = 59), we further assessed whether the main effect of Scenarios was qualified by a significant Age (Young vs. Old; binary variable, Young value was assigned when mean age ≤ 12 months, 29 days) × Scenarios interaction. We did so because most studies on infants' preferences for fair over unfair agents were conducted on older infants (mean age > 12 months, 29 days), while most studies on infants' preference for helpers over hinderers were conducted on younger infants (mean age ≤ 12 months, 29 days). However, the interaction was not significant, QM(2) = 3.01, p = 0.222.

We then asked whether Stimuli influenced infants' choices. Presenting infants with cartoons vs. real events did not influence their socio-moral preferences, QM(1) = 0.13, p = 0.722 (see Figure 3d). Also, the Age (Young, Old) × Stimuli (Cartoon, Real) interaction was not significant, QM(1) = 0.43, p = 0.512; finally, all pairwise comparisons did not reach statistical significance, all ps ≥ 0.490. Moreover, presenting infants with live shows vs.

movies did not affect significantly their preferences, QM(1) = 2.82, $p = 0.093$. Age × Stimuli (Live show, Movies) interaction was not significant, QM(1) = 0.03, $p = 0.869$, and none of the possible pairwise comparisons reached statistical significance, all $ps \geq 0.143$.

Finally, neither Measure nor Agents had a significant influence on infants' choices; QM(1) = 0.17 and QM(2) = 1.28, respectively, all $ps \geq 0.527$.

Discussion

By conducting a meta-analysis on 26 studies which reported a total of 61 effect sizes, we estimated that 68% (CI: 64% to 72%) of infants between 4 and 32 months of age show a preference for a morally good agent (helping, fair or comforting) over a morally bad one (hindering, unfair or hurting). However, we found evidence of a publication bias, as we may have expected given the fact that negative results and replications are often under-reported (Laws, 2013). The publication bias was found also when we included in the analyses the published as well as the unpublished studies that we identified. This suggests that the current meta-analysis, as it is often the case, likely failed to include all the unpublished studies. Then, based on a calculation that adjusted the estimation of the effect size by accounting for the publication bias, we estimated that the true proportion of infants that show a preference for the prosocial agent is 64% (CI: 60% to 69%).

Moreover, we found that effect sizes were influenced by the sample size when two effects from studies with very large samples were included (Cowell & Decety 2015, and Abramson et al. 2016, sample B). Conversely, when these two items were excluded, sample size resulted no longer significant. Although we know that statistical underpowered studies with small sample sizes may threaten the validity of the results, here we found that effect sizes reported by studies with small samples were not significantly different from those reported by studies with large samples. However, we suggest caution in interpreting this result, since almost half of the samples had a size equal to 16 participants, and only three had

more than 32 participants. More research with larger samples (e.g., $N \geq 40$) is needed to properly assess the impact of sample size on the results reported by studies on infants' socio-moral preferences.

Also, given that the distribution of successes in studies with small sample sizes ($N = 16$) is significantly different from the theoretical binomial distribution, we suspect that unpublished effects that may have changed the distribution to lie closer to the theoretical one were not retrieved in the current work. These results are overall consistent with the presence of a publication bias and a file drawer problem; it is likely that conducting studies with small sample sizes (e.g., $N = 16$ or less) and low statistical power (which increases both false positive and false negative results; Fraley & Vazire, 2014; Vadillo, Konstantinidis, & Shanks, 2015), together with the habit of publishing mostly positive results, determines an overestimation of the true effect size. Future research should rely on studies with bigger samples and higher statistical power. Of course, anyone who studies infant psychology knows how difficult and time-consuming it can be. On the other hand, however, high power and big samples increase the accuracy of the effect size estimation as well as the likelihood to detect significant small effects, and thus can prove beneficial also to the researcher's career as she will likely file less studies in the drawer and, in the long run, publish more (Oakes, 2017).

The consistency of the effect

Age did not influence effect sizes. Although more research is needed to strengthen the findings on 2-year-olds (most data has been collected on infants younger than 12 months), we analyzed a sufficiently wide age range to detect possible developmental changes, from the preverbal phase to the subsequent years during which children's social experiences increase in frequency. The result thus suggests that the developmental changes occurring between 4 and 32 months, and the social stimuli to which infants are exposed during this period, do not affect their ability to express socio-moral preferences in manual tasks.

Some of the other potential moderators significantly affected the estimation of the effect size. To begin, we asked whether the fact that the study was conducted by Hamlin and colleagues influenced the estimation. In fact, about half of the effects were reported by Hamlin and colleagues, and, at the same time, some recently published studies failed to replicate them (Cowell & Decety, 2015; Salvadori et al., 2015). Here we showed that, when comparing studies that presented infants with examples of helping/hindering or giving/taking actions, whether the research was conducted by Hamlin and collaborators or by another independent laboratory had a significant impact on the estimated effect size; Hamlin et al. reported on average bigger effect sizes. More studies and replications from independent laboratories are thus needed and would be useful and welcome, especially given the far-reaching theoretical and practical implications of the present findings.

The children's likelihood to prefer the prosocial agent was also influenced by the particular example of prosocial and antisocial behavior presented to the child. After a decade of studies, infants' preferences have been investigated mostly by presenting the child with simple or complex examples of helping and hindering actions (Abramson et al., 2016; Cowell & Decety, 2015; Hamlin, 2013, 2015, 2016; Hamlin et al., 2007; Hamlin et al., 2013; Hamlin & Wynn, 2011; Lee & Song, 2014; Raggio, Hendi, Modesti, Presaghi, & Nicolais, 2015; Salvadori et al., 2015; Steckler et al., 2017; Tasimi & Wynn, 2016; Woo & Hamlin, 2016), fair or unfair distributions (Burns & Sommerville, 2014; Geraci & Surian, 2011; Enright & Sommerville, 2016; Franchin et al., 2015; Surian & Franchin, 2017b), and giving or taking actions (Dahl et al., 2013; Hamlin, 2014; Hamlin et al., 2011; Scola et al., 2015). Analyses showed that presenting children with givers and takers, instead of helpers and hinderers, significantly increases the likelihood of choosing the prosocial agent. Future research should assess whether the ability to express socio-moral preferences in early infancy further extends to other parts of the moral domain. For instance, do infants prefer morally good to morally

bad dominant individuals (Margoni, Baillargeon, & Surian, 2016)?

Other sources of between-studies variability are the dependent variable (whether infants were encouraged to reach for an agent or to help), the class of agents infants were presented with during the manual tasks (puppets, experimenters, or foam shapes), and the modality of stimuli presentation (real events or cartoons). However, none of these factors influenced the estimation of the effect size, indicating that the preference for prosocial agents is consistent across the levels of at least some factors[4]. With respect to the modality of stimulus presentation, results suggest that presenting children with a real event (live show or movie), as opposed to a cartoon, does not facilitate the cognitive processing, as this is reflected by a manual choice.

Interpreting results from socio-moral tasks

Infants' preferences appear to reveal an intuitive moral sense. Still, one can argue that there is no way to decide whether the preference is moral in nature. To make progress, we need to cast further light on the processes that underpin infants' choices. The literature on adult moral judgment, together with the commonsense notion of morality, could help infant researchers to identify a set of characteristics that, when present, make it likely that the evaluation relied on the moral aspects of the events and, at the same time, make it less plausible that it relied on perceptual or simple social cues (Hamlin, 2015b).

First, moral judgments fit within the category of social judgments, and thus they are distinct from non-social judgments such as those on mathematical or inert objects. Infants' socio-moral preferences have been shown to apply only in social contexts that include agents, and not to inanimate objects (Geraci & Surian, 2011; Hamlin et al., 2007; Hamlin & Wynn, 2011). Second, infants' evaluations apply not only to situations in which the child herself was

[4] A limit of these results is that some levels of two variables (Measure and Scenarios) had only a few cases, while other levels included the majority of the cases. E.g., with respect to the variable Measure, 'reach' counted 50 cases, but 'help' only 11 (see Table S1).

helped or hindered, but also to unknown third parties, so we can exclude that considerations related to personal interests trumped the evaluation. Third, infants' evaluations are based on characters' intentions, knowledge state and desires rather than actions outcomes (Hamlin, 2013; Hamlin et al., 2013; Woo, Steckler, Le, & Hamlin, 2017). In this regard, they resemble older children's intent-based moral judgments (Cushman, Sheketoff, Wharton, & Carey, 2013; Killen & Smetana, 2008; Margoni & Surian, 2016, 2017). Finally, infants' evaluations support intuitions about processes of reward and punishment (e.g., Meristo & Surian, 2013, 2014).

Although infants' evaluations possess key elements of adult judgments, we should bear in mind that what we call 'infant socio-moral evaluation' is by no means perfectly isomorphic to older children's and adults' judgment. Infant researchers did not try, so far, to reveal deontic reasoning (Kant, 1785/1959; Kohlberg, 1969), and used measures and tasks different from those used with older children. Moreover, little research has been conducted to connect early abilities with the development of later moral competencies. So, it is at present mainly the fact that infant intuitive preferences possess a certain set of key characteristics of adults' moral evaluations that could justify researchers to consider them 'socio-moral'.

Limitations

Several limitations should be taken into account when interpreting the results of this meta-analysis. As discussed above, most studies had small samples, and almost half of them had exactly $N = 16$. Moreover, only few unpublished studies were retrieved. Finally, almost half of the studies were conducted by a single Lab.

Furthermore, since age was not distributed in a balanced way across the studies and it was confounded with task type, the results concerning its effect should be interpreted with caution. Also, some of the analyses on the influence of moderators Measure or Scenarios are limited by the fact that some levels of the variables had only a few cases. More research with

different stimuli and methodologies could help to assess the generalizability of the present results. A final remark concerns the robustness of infants' preferences. The present meta-analysis can at most clarify whether the preference is robust across individuals, but the question whether the preference is stable within the same individual remains to be tested in future studies (but see Nighbor, Kohn, Normand, & Schlinger, 2017).

Conclusion

The current meta-analysis reported an estimated average proportion equal to 0.68 of infants and toddlers between 4 and 32 months of age that would choose to reach for or help a prosocial agent when it is contrasted with an antisocial agent. We found that children's preferences are consistent across the levels of some factors, showing that the reported effect is generalizable across different tasks and stimuli. We also found that the strength of the preference was not affected by age. This suggests that social input to which infants are exposed and general changes occurring between 4 and 32 months of age do not play a role in shaping the ability to express socio-moral preferences in manual tasks. Finally, we reported evidence of a publication bias and of an unnatural distribution of successes in studies with $N = 16$. So, while this meta-analysis shows that infants' preference for prosocial agents is a well-established fact, it also suggests some caution in interpreting the data from the current literature.

Acknowledgments

We would like to thank XXX XXX and XXX XXX for their helpful comments on an initial version of the manuscript. We also wish to thank the City of XXX, that through a grant to XXX helped the authors to devote themselves to the current work.

References

We indicated with an asterisk (*) the studies that were included in the present meta-analysis. The in-text citations to those studies are instead not preceded by asterisks.

*Abramson, L., Dar, R., Te'eni, T., & Knafo-Noam, A. (2016). Unpublished raw data.

Baillargeon, R., Scott, R., He, Z., Sloane, S., Setoh, P., Jin, K., et al. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer, P. R. Shaver (Eds.), E. Borgida, & J. A. Bargh (Assoc. Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 79-150). Washington, DC: American Psychological Association. https://doi.org/10.1037/14341-003

Bloom, P., & Wynn, K. (2016). What develops in moral development? In D. Barner, & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 347-364). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467630.003.0019

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. (2011). *Introduction to meta-analysis.* Chichester, UK: Wiley.

*Buon, M., Jacob, P., Margules, S., Brunet, I., Dutat, M., Cabrol, D., et al. (2014). Friend or foe? Early social evaluation of human interactions. *PLoS ONE, 9,* e88612. https://doi.org/10.1371/journal.pone.0088612

*Burns, M., & Sommerville, J. (2014). "I pick you": The impact of fairness and race on infants' selection of social partners. *Frontiers in Psychology, 5,* 93. https://doi.org/10.3389/fpsyg.2014.00093

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews: Neuroscience, 14,* 365-376. https://doi.org/10.3410/f.718002370.793478334

Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2016). *Equal praise for equal work: 10-month-olds expect an individual to praise others according to their efforts.* Poster presented at the Biennal International Conference on Infants Studies, New Orleans, LA.

Card, N. A. (2009). *Applied meta-analysis for social science research.* New York: Guilford Publications.

*Cowell, J., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *Proceedings of the National Academy of Sciences, 112,* 12657-12662. https://doi.org/10.1073/pnas.1508832112

Crocetti, E. (2016). Systematic reviews with meta-analysis: Why, when, and how? *Emerging Adulthood, 4,* 3-18. https://doi.org/10.1177/2167696815617076

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29. https://doi.org/10.1177/0956797613504966

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127,* 6-21. https://doi.org/10.1016/j.cognition.2012.11.008

*Dahl, A., Schuck, R., & Campos, J. (2013). Do young toddlers act on their social preferences? *Developmental Psychology, 49,* 1964-1970. https://doi.org/10.1037/a0031460

*Dunfield, K., & Kuhlmeier, V. (2010). Intention-mediated selective helping in infancy. *Psychological Science, 21,* 523-527. https://doi.org/10.1177/0956797610364119

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455-463.

https://doi.org/10.1111/j.0006-341x.2000.00455.x

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634. https://doi.org/10.1136/bmj.315.7109.629

*Enright, E. A., & Sommerville, J. A. (2016). *Selectivity, consistency, and stability of prosociality in infancy and toddlerhood.* Manuscript submitted for publication.

Faraway, J. J. (2006). *Extending the linear model with R.* Boca Raton: Chapman & Hall/CRC.

Ferguson, C., & Brannick, M. (2012). Publication bias in psychological science: Prevalence, methods for indentifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17,* 120-128. https://doi.org/10.1037/a0024445

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS ONE, 9,* e109019. https://doi.org/10.1371/journal.pone.0109019

*Franchin, L., Neira-Gutierrez, I., & Surian, L. (2015). *Do toddlers map the word 'good' to moral qualities?* Unpublished manuscript.

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

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

*Hamlin, K. (2014). Context-dependent social evaluation in 4.5-month-old human infants: The role of domain-general versus domain-specific processes in the development of social evaluation. *Frontiers in Psychology, 5,* 614.

https://doi.org/10.3389/fpsyg.2014.00614

*Hamlin, K. (2015a). The case for social evaluation in preverbal infants: Gazing towards one's goal drives infants' preferences for helpers over hinderers in the hill paradigm. *Frontiers in Psychology, 5,* 1563. https://doi.org/10.3389/fpsyg.2014.01563

Hamlin, K. (2015b). Does the infant possess a moral concept? In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 477-517). Cambridge: MIT Press.

*Hamlin, K. (2016). Unpublished raw data.

Hamlin, K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science, 24,* 589-594. https://doi.org/10.1177/0956797612457785

*Hamlin, K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science, 16,* 209-226. https://doi.org/10.1111/desc.12017

*Hamlin, K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26,* 30-39. https://doi.org/10.1016/j.cogdev.2010.09.001

*Hamlin, K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557-560. https://doi.org/10.1038/nature06288

Hamlin, K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluation. *Developmental Science, 13,* 923-929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

*Hamlin, K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences, 108,* 19931-19936. https://doi.org/10.1073/pnas.1110306108

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency

in meta-analysis. *British Medical Journal, 327,* 557-560. https://doi.org/10.1136/bmj.327.7414.557

Holvoet, C., Scola, C., Arciszewski, T., & Picard, D. (2016). Infants' preference for prosocial behaviors: A literature review. *Infant Behavior and Development, 45,* 125-139. https://doi.org/10.1016/j.infbeh.2016.10.008

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2,* e124. https://doi.org/10.1371/journal.pmed.0020124

Kanakogi, Y., Okumura, Y., Inoue, Y., Kitazaki, M., & Itakura, S. (2013). Rudimentary sympathy in preverbal infants: Preference for others in distress. *PLoS ONE, 8,* e65292. https://doi.org/10.1371/journal.pone.0065292

Kant, I. (1785/1959). *Foundation of the metaphysics of morals.* Indianapolis: Bobbs-Merrill.

Killen, M., & Smetana, J. (2008). Moral judgment and moral neuroscience: Intersections, definitions, and issues. *Child Development Perspectives, 2,* 1-6. https://doi.org/10.1111/j.1750-8606.2008.00033.x

Killen, M., & Smetana, J. G. (2015). Origins and development of morality. In M. E. Lamb (Ed.), *Handbook of child psychology and developmental science, Vol. 3* (pp. 701-749). New York: Wiley-Blackwell.

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

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14,* 402-408. https://doi.org/10.1111/1467--9280.01454

Laws, K. R. (2013). *Negativland – a home for all findings in psychology.* BMC Psychology, 1, 1-8. https://doi.org/10.1186/2050-7283-1-2

*Lee, Y., & Song, H. (2014, July). *Six-month-olds' sensitivity to the intentions of helpful behaviors.* Poster presented at the International Conference on Infants Studies, Berlin, Germany.

Lee, Y., Yun, J. E., Kim, E. Y., & Song, H. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLoS One, 10,* e0135588. https://doi.org/10.1371/journal.pone.0135588

Lobo, M. A., & Galloway, J. C. (2013). The onset of reaching significantly impacts how infants explore both objects and their bodies. *Infants Behavior and Development, 36,* 14-24. https://doi.org/10.1016/j.infbeh.2012.09.003

Margoni, F., Baillargeon, R., & Surian, L. (2016). *Infants distinguish between leaders and bullies.* Manuscript submitted.

Margoni, F., & Surian, L. (2016). Explaining the U-shaped development of intent-based moral judgments. *Frontiers in Psychology, 7,* 219. https://doi.org/10.3389/fpsyg.2016.00219

Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development, 41,* 46-64. https://doi.org/10.1016/j.cogdev.2016.12.001

McDonnell, P. M. (1975). The development of visually guided reaching. *Perception & Psychophysics, 18,* 181-185. https://doi.org/10.3758/bf03205963

Meristo, M., Strid, K., & Surian, L. (2016). Preverbal infants' ability to encode the outcome of distributive actions. *Infancy, 21,* 353-372. https://doi.org/10.1111/infa.12124

Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition, 129,* 102-113. https://doi.org/10.1016/j.cognition.2013.06.006

Meristo, M., & Surian, L. (2014). Infants distinguish antisocial actions directed towards fair and unfair agents. *PLoS ONE, 9,* e110553.

https://doi.org/10.1371/journal.pone.0110553

Nighbor, T., Kohn, C., Normand, M., & Schlinger, H. (2017). Stability of infants' preference for prosocial others: Implications for research based on single-choice paradigms. *PLoS ONE, 12,* e0178818. https://doi.org/10.1371/journal.pone.0178818

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infants looking-time research. *Infancy, 22,* 436-469. https://doi.org/10.1111/infa.12186

*Raggio, E., Hendi, S. F., Modesti, C., Presaghi, F., & Nicolais, G. (2015). Temperament and attachment in the development of moral precursors: Preliminary data. *Infanzia e Adolescenza, 14,* 197-217.

*Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., et al. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PLoS ONE, 10,* e0140570. https://doi.org/10.1371/journal.pone.0140570

*Scarf, D., Imuta, K., Colombo, M., & Hayne, H. (2012). Social evaluation or simple association? Simple associations may explain moral reasoning in infants. *PLoS ONE, 7,* e42698. https://doi.org/10.1371/journal.pone.0042698

Schmidt, M., & Sommerville, J. (2011). Fairness expectations and altruistic sharing in 15-month-old infants. *PLoS ONE, 6,* e23223. https://doi.org/10.1371/journal.pone.0023223

*Scola, C., Holvoet, C., Arciszewski, T., & Picard, D. (2015). Further evidence for infants' preference for prosocial over antisocial behaviors. *Infancy, 20,* 684-692. https://doi.org/10.1111/infa.12095

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23,* 196-204. https://doi.org/10.1177/0956797611422072

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

olds fail to generate social evaluations of individuals who behave inconsistently. *Cognition, 167,* 255-265. https://doi.org/10.1016/j.cognition.2017.03.018

Surian, L., & Franchin, L. (2017a). Infants reason about deserving agents: a test with distributive actions. *Cognitive Development, 44*, 49-56. http://dx.doi.org/10.1016/j.cogved.2017.08.009

*Surian, L., & Franchin, L. (2017b). Toddlers selectively help fair agents. *Frontiers in Psychology, 8*, 944. https://doi.org/10.3389/fpsyg.2017.00944

*Tasimi, A., & Wynn, K. (2016). *Infants have enduring memory of who's nice and who's mean.* Unpublished manuscript.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J., et al. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development, 64,* 1058-1098. https://doi.org/10.2307/1131327

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23,* 87-102. https://doi.org/10.3758/s13423-015-0892-6

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48. https://doi.org/10.18637/jss.v036.i03

von Hofsten, C. (1984). Developmental changes in the organization of prereaching movements. *Developmental Psychology, 20,* 378-388. https://doi.org/10.1037//0012-1649.20.3.378

*Woo, B. M., & Hamlin, K. (2016). Unpublished raw data.

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

Ziv, T., & Sommerville, J. (2016). Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Development.* Advance online publication. https://doi.org/10.1111/cdev.12674