Not so innocent: Reasoning about costs, competence, and culpability in very early childhood

Julian Jara-Ettinger (jjara@mit.edu)
Joshua B. Tenenbaum (jbt@mit.edu)
Laura E. Schulz (lschulz@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139 USA

Abstract

Social evaluations depend on our ability to interpret other people's behavior. In adults, these evaluations are influenced by our perception of the competence and motivation of the agent: helping when it is difficult to help is praiseworthy; not helping when it is easy to help is reprehensible. Here we look at young children's capacity to make competence attributions and its relation to their social evaluations. We find that as early as 18-months, infants can use the time and effort associated with achieving a goal-directed action to distinguish agents, and that infants prefer more competent agents. When asked to choose between two agents who act as moral bystanders and refuse to engage in a helpful action, we find a sustained preference for the more competent agent until the age of three, when the preference is reversed. We argue that the ability to calculate the cost and benefits of goal-directed action originates in early childhood and plays a fundamental role in moral reasoning.

Keywords: Action Understanding; Morality; Social Cognition; Theory of Mind.

Introduction

The past decade has seen a revolution in our understanding of psychosocial reasoning in early childhood. Recent findings suggest that infants infer the false beliefs of others (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Kovács, Téglás, & Endress, 2010), distinguish helpers, hinderers, and moral bystanders (Kuhlmeier, Wynn, & Bloom, 2003; Hamlin, Wynn, & Bloom, 2007); draw different inferences about actions directed towards member of in-groups and out-groups (Baillargeon et al., in press); predict actions based on social dominance (Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011), judge third party agents transitively, based on how they interact with moral transgressors (Hamlin, Wynn, Bloom, & Mahajan, 2011; Sloane, Baillargeon, & Premack, 2012); and consider agents' knowledge about a target agent's preferences in making moral judgments (Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013). The discovery of infants' sophisticated social intelligence is among the most exciting recent developments in the field of cognitive science. However, to date relatively little is understood about the types of computations that underlie these social judgments.

Rational Planning, Social Evaluations, and a Naïve Utility Calculus

Here we propose a new approach to thinking about social reasoning in infancy, drawing on the insight that the ability to reason about goal-directed action is at the core of our cognition about agents. (See Carey, 2009; Gergely & Csibra, 2003 for review). Consistent with a large body of prior work, we assume that inferences about agents' goal-directed actions are governed by a principle of rational expectation: the idea that agents act efficiently to achieve their goals (e.g., Scott & Baillargeon, 2013; Gergely & Csibra, 2003). Computational work on the principle of rational expectation as probabilistic inference over rational planning has been used to successfully model adults' reasoning about agents' goals (Baker, Saxe, & Tenenbaum, 2009, 2011; Ullman et al., 2010; Jara-Ettinger, Baker, & Tenenbaum, 2012).

The principle of rational expectation is predicated on the understanding that agents act in ways that will minimize costs and maximize rewards. We propose that the ability to compute the costs and benefits of actions forms the heart of a naïve utility calculus that supports inference at the earliest stages of children's theories of agency. (See Jara-Ettinger, Gweon, Tenenbaum, & Schulz in prep, for a detailed version of this argument and experimental studies in childhood). Here we provide an informal description of this approach and test one of its qualitative predictions: that an analysis of the cost functions associated with agent actions is central to the moral judgments even of very young children.

Intuitively, adult social evaluations are influenced by our perception of how much an action will cost the agent who performs it. Imagine for instance, that your neighbor, Sally, watches a child struggle to reach a package on the top shelf of a grocery store. Sally stands by and does nothing at all. Although there is no intrinsic relationship between height and moral worth, you may well judge Sally less harshly if she is 4'11" than if she is an NCAA Division 1 basketball player.

What analysis underlies this inference? We suggest that in evaluating and predicting agents' actions, observers automatically compute the cost of actions. The perceived cost of an action (controlling for constraints imposed by the environment) reflects inferences about the agents' level of competence; the perceived benefits of the action to the agent reflect inferences about the agents' level of motivation. Motivation and competence jointly affect the probability of the agents' actions so the two attributions trade-off with each other. If we know that an agent is highly motivated and she fails to act, we may infer that she is incompetent; conversely, if we know the agent is highly competent and she fails to act, we may infer that she is unmotivated. Morally, lack of competence to help is an exonerating factor; lack of motivation is not.

More generally social evaluations depend heavily on the agent's motivation (Cushman, 2008; Knobe, 2005; Young, Cushman, Hauser, & Saxe, 2007). This may not be known and must then be inferred. If the agent performs a morally worthy action, then the higher our estimate of the cost of the action, the higher our estimate of the agents' motivation to act morally. Similarly, if the agent fails to act, then the lower our estimate of the cost of action, the lower our estimate of the agent's motivation. Ambiguity arises when the agent acts but the cost of action is very low, or the agent fails to act but the cost of action is very high; in such cases, we may be unsure of the agent's level of motivation. In our example, if Sally is 4'11", there is a high cost to reaching the shelf. This renders her failure to act ambiguous. Did she not want to help or was it simply too hard for her to do so? By contrast, if Sally is an NCAA basketball player, we can infer that the cost of reaching a shelf is low; thus we are more confident that her failure to act derives from a morally suspect lack of motivation.

We propose that these kinds of considerations are part of a general calculation of a cost function that, even early in development, is used to reason about goal-directed behavior and interpret agents' actions. However, to date no empirical work has looked at how differences in the cost function of agent actions affect children's evaluative and moral judgments. Similarly, no previous computational work has looked at how learners might compute the cost function of agent actions; work on goal inference has implicitly assumed that the cost function of actions is known (e.g., Baker et al., 2009; Ullman et al., 2010).

Here we test the prediction that very young children can estimate the cost functions associated with agents' actions and that this analysis affects children's moral judgments. In Experiment 1, we test the basic premise that children can use the perceived cost of actions to estimate agents' competence. We predict that at baseline children will prefer more (versus less) competent agents. In Experiment 2, we look at whether children can use differences in the cost of actions to infer differences in agents' motivations. We predict that when agents are moral bystanders, children may overcome their baseline preference for competent agents and be more likely to consider the merits of incompetent (but potentially more well-intended) agents.

Experiment 1: Early Competence Attribution

In Experiment 1 we look at whether toddlers can use the time and effort associated with achieving a goal-directed action to estimate the cost of the action to the agents. We also look at whether toddlers have an early preference for competent agents.

Participants

Twenty-four toddlers (mean age (SD): 21.19 months (97 days), range 16.8-28 months, 16 males) were tested at an urban children's museum[1]. Five children were excluded from analysis: four by decision of a blind coder and one for parental interference (See Results). All subjects were tested at an urban children's museum.

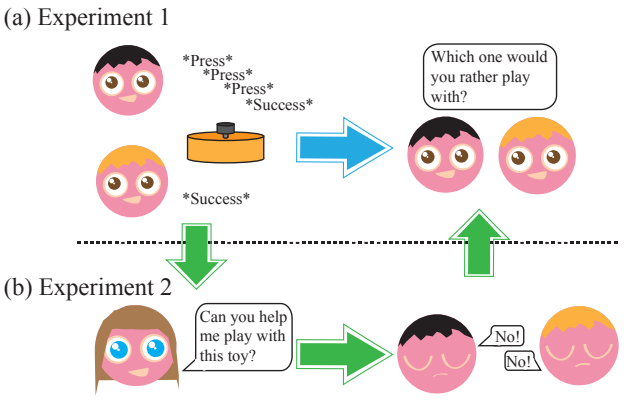

Figure 1: Procedure for Experiments 1 and 2. Both experiments begin by introducing two puppets and a toy. One puppet (the Competent agent) was able to make the toy play music on the first attempt; the other puppet (the Incompetent agent) succeeded only after many attempts. In Experiment 1 (blue arrow), children were then asked to choose one of the puppets to play with. In Experiment 2 (green arrows), after the child saw both puppets activate the toy, the parent turned around and asked each puppet for help with the toy. Both puppets refused. As in Experiment 1, children were then asked to choose one of the puppets to play with.

Stimuli

Participants were shown two puppets and a yellow cylindrical toy with a black button at the top. The toy played music when the button was pressed.

Procedure

Participants were tested in a quiet room at the museum. The child's parent was seated on a chair facing away from the testing table and the parent was asked to hold the toddler over his or her shoulder. Thus the child could see the stimuli but the parent could not.

Once the parent and toddler were positioned the experimenter presented the yellow toy to the child and introduced the two puppets. See Figure 1. He said, "Here are my two friends! They are going to show you how the toy works." Both puppets were continuously present throughout the experiment and each puppet approached the toy (order counterbalanced between participants) one at a time. The puppet said, "It's my turn!" and then pressed the button. When the toy activated, the toy played a song for approximately 10 sec-

[1] 12 additional toddlers were recruited but never included in the study because they declined to participate in a warm-up task, in which the child was asked to choose between two stuffed elephants.

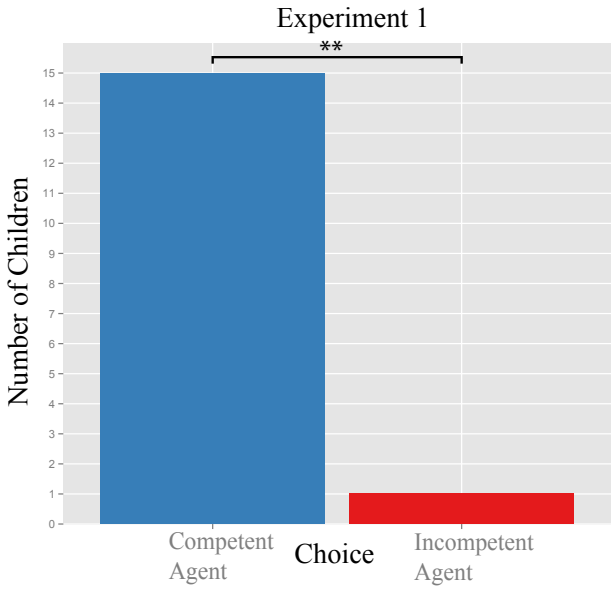

Figure 2: Results from Experiment 1: Number of children choosing each agent. $** = p < .001$ by binomial test.

onds and then the puppet released the button. During this time, both puppets moved rythmically to the sound of the song. After releasing the button, each puppet said "Yay!" to celebrate the success.

The puppets differed in how many attempts it took them to activate the toy. The competent agent was always able to make the toy play music on the first attempt. The incompetent puppet tried several times to activate the toy (flattening his hand over the button but not depressing it fully). After the third or fourth failed attempt, the incompetent puppet backed away to look at the button, and then tried again. The incompetent puppet made a few more failed attempts and then successfully activated the toy. (The number of total attempts ranged from 6 - 8 trials across participants, allowing some flexibility in maintaining the child's attention to the task.) After the show, the parent was asked to turn around and to place their child at a marker on the middle of the edge of a lower table. The experimenter placed both puppets on opposite sides of the table equidistant from the child and asked the child which one she wanted to play with.

Results and Discussion

All videotapes were coded by a coder blind to condition. Four children were excluded from analysis due to the coders' judgment that the puppets were not placed equidistant from the child. One additional child was excluded from analysis due to parental interference. The coder recorded the toddlers' first contact with a puppet following the prompt. If the child did not make a choice within a 30-second window following the prompt, the experiment was ended. Three children did not make a choice. Of the 16 children who did make a choice, 15 preferred the competent agent ($p < 0.001$ by binomial test). See Figure 2.

In our design, the incompetent agent both made more attempts to activate the toy and took longer to activate the toy. Additionally, after some initial failures, the incompetent puppet studied the toy before trying again. Thus there were redundant cues to the agent's incompetence and we do not know whether toddlers' preferences were driven by the overall effort to achieve the goal, the time to achieve the goal (and thus perhaps the relative novelty of the puppet who achieved the goal more quickly), or a more abstract judgment about these factors as indices of competence per se. Future research might look at the range of factors that affect toddlers' inferences about the cost of agent actions. However, the result of Experiment 1 give strong evidence that by 18 months, children distinguish agents from differential cues to competence and prefer agents who appear to incur fewer costs to achieve a goal.

Experiment 2: Competence and Social Evaluations

In Experiment 2, we look at how children's judgment of agent competence affects their social evaluation. Because pilot work suggested that the task in Experiment 2 was more demanding than the one in Experiment 1, we tested slightly older children: two and three-year-olds.

Participants

Seventeen two-year-olds (mean age (SD): 30.8 months (83 days), range 26.6-34.9 months, 9 males); one was dropped from analysis for failure to make a choice. Thirty three-year-olds (mean age (SD): 42 months (104 days), range 36-50.09 months, 17 males) were recruited in the test condition; 7 were dropped from analysis, 4 by decision of a blind coder and 3 for failures to make a choice. An additional 9 three-year-olds (mean age (SD): 35.4 months (131.67 days), range 29.1-42.03 months, 4 males) were recruited for a control condition, 1 was dropped from analysis due to failure to make a choice. All subjects were tested at an urban children's museum[2].

Stimuli

The stimuli used in Experiment 2 were identical to stimuli used in Experiment 1.

Procedure

The protocol in Experiment 1 was identical to the protocol in Experiment 2 with the following exceptions (See Figure 1). Because the children were older, they were given a choice of sitting in a small chair or standing in front of the testing table, behind the parent's chair. Additionally, before the experiment began, the parents were given a script to read telling them that when prompted to do so, they should turn around and pick up

[2]4 additional two-year-olds and 3 three-year-olds were recruited but never included in the study because they declined to participate in a warm-up task, in which the child was asked to choose between two stuffed elephants.

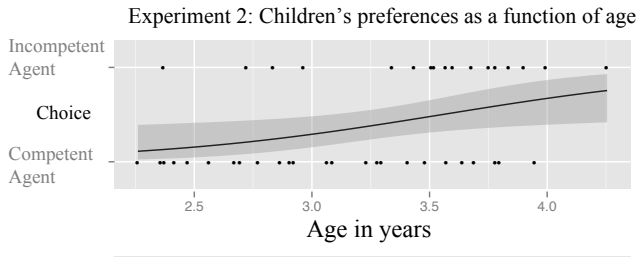

Figure 3: Children's choice of puppet in the test condition of Experiment 2 as a function of their age. The logistic regression with 95% confidence interval is shown on top. At early ages we find a preference for the competent agent, which disappears in older subjects.

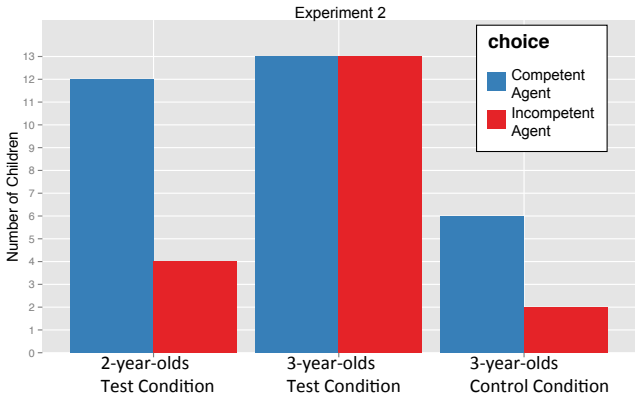

Figure 4: Number of subjects choosing to play with the competent (red) and incompetent agent (blue) in each age and condition of Experiment 2. $* = p < .05$.

the toy from the table. The experimenter would then place one puppet at a time in front of them. Parents in the test condition were instructed to ask each puppet: "Can you help me make the toy go?" Parents in the control condition were instructed to ask each puppet "Do you have a toy like this at home?"

Otherwise, the first part of the protocol proceeded as in Experiment 1. After both the competent and incompetent puppets successfully made the toy play music, but before the child was asked to make a choice, both puppets were removed and the toy was placed in the middle of the table. At this point the parent was asked to turn around. The parent picked up the toy and the experimenter returned a puppet to the middle of the table (order of puppets counterbalanced). Only one puppet was visible at a time. After the parent asked the puppet the target question, the puppet looked at the toy, then at the parent and said "No!" The puppet then turned around and hid under the table. This was repeated with the next puppet. To ensure that the child understood, in the test condition the experimenter said, "No one seems to want to help!" In the control condition he said, "No one seems to have this toy!" The questions and answers were then repeated with each puppet a second time.

After each puppet had said "no" twice, the experimenter took the toy from the parent and asked the child to stand on a marker in the center of a table edge. As in Experiment 1, the experimenter then set each puppet on opposite sides of the table, equidistant from the child and asked the child which puppet she would rather play with.

Results and Discussion

Results were coded from videotape by a coder blind to conditions, as in Experiment 1. Children were excluded from analysis if, in the coder's judgment, the puppets were not placed equidistant from the child or if children did not make a choice within the 30-second window, resulting in 16 2-year-olds and 23 3-year-olds in the test condition and 8 3-year-olds in the control condition (See Participants).

In the test condition, a logistic regression showed an effect of age on children's preferences: older children were more likely than younger children to choose the incompetent puppet ($p < 0.02$). See Figure 3.

We followed-up with planned comparisons of the two and three-year-olds separately. See Figure 4. The two-year-olds showed a robust preference for the competent puppet. Of the 16 two-year-olds who made a choice, 12 chose the competent puppet ($p < .05$ by binomial test). By contrast, the three-year-olds in the test condition chose between the puppets at chance; 13 of the 26 three-year-olds chose the competent puppet ($p = ns$ by binomial test).

These results are consistent with the possibility that three-year-olds can use differences in agents' competence to attribute differences in agents' motivation, and can overcome their baseline preference for competent agents if agents fail to act helpfully. Arguably however, the three-year-olds chose at chance because they simply forgot which puppet was more competent (perhaps because the three-year-olds were more engaged than the two-year-olds by the puppets' refusals).

To see whether three-year-olds retained the competence information we looked at three-year-olds' performance in the control conditions. Failure to recall the more competent puppet seems unlikely to explain the results; preliminary results from the control condition suggest that the three-year-olds have no difficulty remembering which puppet was more competent when moral culpability is not at issue: 6 of the 8 three-year-olds showed a preference for the more competent puppet.

These results suggest that by the age of three, children can override a preference for competent agents if those agents act as moral bystanders. Given that it is morally objectionable to refrain from helping when a helpful action is relatively low cost, three-year-olds seem to be able to look more favorably on agents who have the excuse of incompetence to exonerate them.

The current findings are consistent with previous work sug-

gesting that both toddlers (Tomasello et al., 2005) and chimpanzees (Call, Hare, Carpenter, & Tomasello, 2004) differentiate between agents who are unwilling to act helpfully from those who are unable to act helpfully. Here we also find that children distinguish competence to help from motivation to help. Critically however, participants in the earlier studies could assess the agents' motivation directly, from overt behavioral cues: serious, if failed, attempts to help indicated a motivated agent; "teasing" indicated an unmotivated one. Additionally, the "unable" agent in the earlier studies was genuinely unable: every attempt the agent made failed.

By contrast, in the current study, both agents were unwilling to help (both puppet said "no" and turned away from the parent) and neither agent was unable to help (both puppets were in fact able to activate the toy). Children could only evaluate the agents on the basis of graded differences in the agents' competence; however, this is precisely the kind of ability children should have if, as we have proposed, social reasoning in early childhood is informed by a naïve utility calculus, supporting computations of the costs and benefits of actions.

General Discussion

Consistent with the idea that a naïve utility calculus is integral to children's understanding of agents, we found that inferences about the relative cost of agents' actions affect children's social evaluations from very early in development. Toddlers seem to be sensitive to cues associated with the relative competence of agents and prefer agents who achieve goals quickly and easily to agents who achieve the same goals at apparently higher costs. By the age of three, children seem to be able to use differences in agent competence as grounds for evaluating agents differently, even when the agents act identically in refusing to act at all.

As noted, we provided redundant cues to the competence and incompetence of these agents, including the time it took for each agent to make the toy play music, and the number of times each puppet pressed the button. We do not know to what extent toddlers' preferences were driven by each individual cue, or if their choice was guided by a more abstract representation of competence. Future research can shed light on the full range of cues we use to infer an agent's competence both in the physical domain and the epistemological domain, where some form of competence preference has also been found (Koenig, Clément, & Harris, 2004; Pasquini, Corriveau, Koenig, Harris, et al., 2007).

There are several hypotheses consistent with the developmental change we observed between the age of two and three. One possibility is that toddlers distinguish competent and incompetent agents, but they do not infer that relative competence implies an obligation to act helpfully, or that relative incompetence exonerates an agent from such actions. A related possibility is that toddlers might make categorical distinctions between classes of behavior (e.g., "helping", "not helping", and "hindering") but make no distinctions within each category; because both puppets in our paradigm refused to help, toddlers might find them equally blameworthy. A final intriguing possibility is that both two and three-year-olds can integrate judgments of agents' competence with moral judgments, but the children find themselves in a moral dilemma (and resolve the dilemma differently at different ages): they believe the incompetent agent is less culpable; however, they also believe it is a good idea to affiliate with competent agents. Future work is necessary to disambiguate these possibilities.

Additionally, because the children were given a forced choice between two agents, we do not know whether the one and two-year-olds' choices were based on a preference for the competent agent, an aversion to the incompetent one, or both. Similarly, we do not know whether the three-year-olds' choices reflect a relatively greater preference for the less competent (and therefore morally exonerated) agent, a relative devaluing of the more competent (and therefore morally culpable) agent, or both. Further research might disambiguate the specific attributions underlying children's preferences.

What this study does show is that human beings are sensitive to the cost of actions very early in development and form an early preference for competent agents. As children progress through early childhood, they become increasingly able to use inferences about an agent's competence to draw inferences about the agent's moral status. At an age when children themselves are still largely both incompetent and innocent, their ability to understand how the one characteristic might bear upon the other suggests remarkably sophisticated inferential abilities and highlights the importance of building a new theoretical synthesis for understanding the development social reasoning.

Acknowledgments

We thank the Boston Children's Museum and the families who volunteered to participate. Thanks to Salvador Esparza, Eric Garr, and Aviana Polsky for help with data collection. We thank Rachel Magid for help with coding the data. Special thanks to Hyowon Gweon and Kim Scott for helpful comments and discussions. This research was funded by the National Science Foundation and the Simons Center for the Social Brain.

References

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., et al. (in press). Psychological and sociomoral reasoning in infancy. *APA Handbook of Personality and Social Psychology: Vol. 1. Atittudes and Social Cognition.*.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Thirtieth Third Annual Conference of the Cognitive Science Society* (p. 2469-2474).

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). unwillingversus unable: chimpanzees understanding of human intentional action. *Developmental science*, *7*(4), 488–498.

Carey, S. (2009). *The origin of concepts*. Oxford University Press, USA.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Developmental science*, *16*(2), 209–226.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557-560.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011, December). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19931–6.

Jara-Ettinger, J., Baker, C. L., & Tenenbaum, J. B. (2012). Learning what is where from social observations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 515–520).

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (in prep). You can't always get what you want: Inferring competence from preference in early childhood.

Knobe, J. (2005). Theory of mind and moral cognition: exploring the connections. *Trends in Cognitive Sciences*, *9*(8), 357 - 359.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Kovács, A. M., Téglás, E., & Endress, A. D. (2010, December). The social sense: susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, *330*(6012), 1830–4.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003, September). Attribution of Dispositional States by 12-Month-Olds. *Psychological Science*, *14*(5), 402–408.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Pasquini, E. S., Corriveau, K. H., Koenig, M., Harris, P. L., et al. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental psychology*, *43*(5), 1216–1226.

Scott, R. M., & Baillargeon, R. (2013). Do infants really expect agents to act efficiently?: A critical test of the rationality principle. *Psychological Science*.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological science*, *23*(2), 196–204.

Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation through Attribution of False Belief By Two-Year-Olds. *Psychological Science*, *18*, 587–592.

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011, January). Big and mighty: preverbal infants mentally represent social dominance. *Science (New York, N.Y.)*, *331*(6016), 477–80.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H., et al. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–690.

Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2010). Help or hinder: Bayesian models of social goal inference. In *Advances in Neural Information Processing Systems 22* (pp. 1874–1882).

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235-8240.