CHILD DEVELOPMENT

Child Development, xxxx 2019, Volume 00, Number 0, Pages 1–15

Social Pragmatics: Preschoolers Rely on Commonsense Psychology to Resolve Referential Underspecification

Julian Jara-Ettinger
Yale University

Sammy Floyd
Princeton University

Holly Huey
UCSD

Joshua B. Tenenbaum and Laura E. Schulz
MIT

Four experiments show that 4- and 5-year-olds (total N = 112) can identify the referent of underdetermined utterances through their Naïve Utility Calculus—an intuitive theory of people's behavior structured around an assumption that agents maximize utilities. In Experiments 1–2, a puppet asked for help without specifying to whom she was talking ("Can you help me?"). In Experiments 3–4, a puppet asked the child to pass an object without specifying what she wanted ("Can you pass me that one?"). Children's responses suggest that they considered cost trade-offs between the members in the interaction. These findings add to a body of work showing that reference resolution is informed by commonsense psychology from early in childhood.

When we talk to others, the meaning of what we say is often revealed by the context in which we say it. Suppose you offer coffee to a friend at 11 p.m. and she says "coffee would keep me awake." This answer probably means that she doesn't want any. But if you were working toward a deadline, the same answer might very well mean that she does (Wilson & Sperber, 2002). Although it is tempting to credit the listener for successfully inferring the speaker's intended answer, these kinds of interactions can only be successful when the speaker provides enough information in the first place (Clark & Bangerter, 2004; Clark & Schaefer, 1987; Clark, Schreuder, & Buttrick, 1983; Clark & Wilkes-Gibbs, 1986; Grice, 1975; see Graf & Davies, 2014 for review). As speakers and listeners, we can only do this effectively by relying on our commonsense psychology—our general expectations about how other people's mental states relate to their actions (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016): Speakers must decide how much to say so listeners can recover the intended message, and listeners must infer what meaning is justified given what the speaker said.

Consistent with this, developmental research has shown that the interaction between commonsense psychology and language understanding is at work from early in life. By age 2, children are already sensitive to other people's knowledge (Bohn, Zimmermann, Call, & Tomasello, 2018; O'Neill, 1996; Southgate, Chevallier, & Csibra, 2010), past experience (Saylor & Ganea, 2007; Saylor, Ganea, & Vázquez, 2011), visual perspective (Baldwin, 1991; Koenig & Echols, 2003; Vaish, Demir, & Baldwin, 2011), and information in common ground (Akhtar, Carpenter, & Tomasello, 1996; Liebal, Carpenter, & Tomasello, 2010; Moll & Tomasello, 2006; see Bohn & Köymen, 2018 for review), although these capacities continue to develop throughout early childhood (Moll, Carpenter, & Tomasello, 2007; Morisseau, Davies, & Matthews, 2013; Nilsen & Graham, 2009). These studies show that the building blocks necessary for reference resolution are in place in early childhood. However, in some cases an even more fine-grained understanding of the structure of the event is required. Compare, for instance,

We thank the Boston Children's Museum and the families who volunteered to participate. We thank Nate Clark, Mary DePascale, Madeline Klein, and Caiqin Zhou for help with recruitment, piloting, and data collection. We thank Paula Rubio-Fernández, Tracy Reuter and five anonymous reviewers for helpful comments on an earlier version of this manuscript. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF-STC award CCF-1231216, and by Simons Center for the Social Brain (SCSB).

Correspondence concerning this article should be addressed to Julian Jara-Ettinger, Department of Psychology, Yale University, 2 Hillhouse Ave, New Haven, CT 06520. Electronic mail may be sent to julian.jara-ettinger@yale.edu.

DOI: 10.1111/cdev.13290

1. Anna should pass the salt to Claire because _she_ is close to it.
2. Anna should pass the salt to Claire because _she_ is far away from it.

Although the *she* pronouns could apply to Anna or to Claire, we readily assign them to different referents in the two sentences. This is because it only makes sense for someone to pass the salt if they are closer to it than the person who wants it. Here, the resolution of the ambiguity does not depend on Anna's and Claire's mental states but on the relative costs and benefits each of them would incur from the action. Past research on reference resolution has focused primarily on situations where the speaker or listener's knowledge is manipulated. In contrast, less is known about how reference resolution might be facilitated by an understanding of others' costs and rewards. We propose that these commonsense inferences are supported by people's Naïve Utility Calculus—and intuitive theory of action understanding that works through the assumption that agents act to maximize their subjective utilities (Jara-Ettinger et al., 2016; Liu, Ullman, Tenenbaum, & Spelke, 2017; Lucas et al., 2014). We begin by reviewing the Naïve Utility Calculus, and then turn to the role it may play in reference resolution.

The Naïve Utility Calculus

Research suggests that commonsense psychology is structured around the assumption that agents act by maximizing their subjective utilities—the difference between the costs they incur and the rewards they obtain (Jara-Ettinger et al., 2016; Jern, Lucas, & Kemp, 2017; Liu et al., 2017; Lucas et al, 2014). Most directly, this assumption implies that agents will pursue goals efficiently (the smaller the cost, the higher the utility; Gergely & Csibra, 2003) and only when the rewards outweigh the costs (otherwise, not doing anything at all yields a higher utility). This assumption also enables observers to parse other people's actions into judgments about the underlying costs they expected to incur and the rewards that they expected to obtain. For instance, if your friend walked past the cafe around the corner and kept walking to a distant coffee shop, she clearly prefers their coffee. The Naïve Utility Calculus predicts this inference about her preference, as incurring a higher cost is only justified by the presence of a higher reward. By contrast, if your friend got her coffee around the corner, you would not necessarily assume that your friend thinks the nearby coffee shop is superior to the distant one. The Naïve Utility Calculus predicts this as well, as the coffee nearby may have a high utility because the cost getting there is low and not because the reward is particularly high.

Even young children's intuitions about other people's preferences align with the predictions of the Naïve Utility Calculus. Five-year-olds infer high rewards when an agent incurs a high cost and, conversely, they infer high costs when an agent foregoes a high reward (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015; Pesowski, Denison, & Friedman, 2016). Children at this age also understand that agents do not maximize the utilities they obtain but the utilities they *expect* to obtain. For instance, children understand that an agent who has never tried a set of fruits before will only choose what she expects to like best and that she may not necessarily like it in the end. Similarly, when agents incur a high cost to obtain a reward, children only infer a strong preference if the agent was aware of the costs when making her choice (Jara-Ettinger, Floyd, Tenenbaum, & Schulz, 2017). At an even earlier age, 2-year-olds condemn agents who refuse to incur a low cost to help another agent relative to agents who refuse to incur a high cost to help (Jara-Ettinger, Tenenbaum, & Schulz, 2015).

The previous studies focus on the interpretation of others' goal-directed actions. However, the Naïve Utility Calculus also makes predictions about how we interpret requests. To illustrate this, consider a simple request for help (Figure 1). For someone to agree to help, their motivation to help (their reward) must outweigh the cost of helping. This means that the easier it is for them to help, the more likely that they will agree. Thus, when more than one person can help, the Naïve Utility Calculus predicts that people should ask whomever can help more easily. This is true for two reasons. First, as described above, people who can help easily are more likely to agree to help (Figure 1b). Second, if we care about others, we should also prefer to minimize their costs. Thus, the Naïve Utility Calculus predicts that when a request for help is ambiguous (e.g., "Can _you_ help me?" in the presence of various agents), people should infer that the request is directed at the agent who incurs fewest costs. We test this prediction in Experiments 1–2.

By a similar logic, the Naïve Utility Calculus predicts that we should be more likely to ask for help with things that are harder to achieve alone and easier to achieve for the helper. To illustrate this, suppose you ask a friend who is across the room to

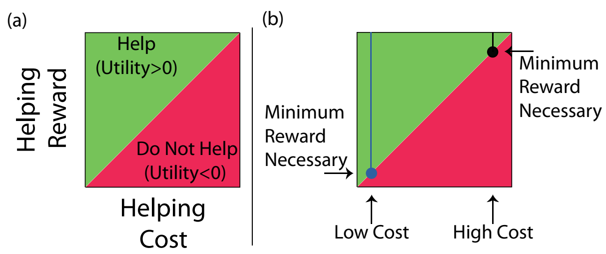

Figure 1. (a) Qualitative visualization of the utility as a function of the costs and the rewards. The green area shows regions where the utility is positive (rewards outweigh costs) and the red area shows regions where the utility is negative (costs outweigh rewards). (b) Range of rewards that motivate an agent to act when the cost is low and when the cost is high. Agents who can help more easily need a lower reward for agreeing to help.

pass you a salt shaker that is within your arm's reach. Clearly, the cost you are asking your friend to incur is higher than the cost you would have to incur if you chose to get the salt shaker yourself. This unambiguously reveals that you care less about your friend's costs relative to your own. Your friend, inferring this, should reasonably refuse to help and perhaps even consider you rude. If, instead, the salt shaker was close to your friend and far away from you, asking them to pass it over would now be reasonable because the cost that you ask your friend to incur is lower than the one you would have to incur yourself. Thus, the Naïve Utility Calculus predicts that we should expect people to only ask for help when the cost of doing something for themselves is higher than the cost that the helper would need to incur. Given an ambiguous request ("Can you pass me *that* one?"), we predict that children will resolve the referent by inferring that it involves the request that is less costly for the target than the speaker. We test this prediction in Experiments 3–4.

The Present Study

Here we test if children can resolve referential ambiguity through their Naïve Utility Calculus. We focus on the earliest ages in which children have a mature Naïve Utility Calculus: 4- and 5-year-olds (Jara-Ettinger et al., 2016; Jara-Ettinger, Gweon, et al., 2015). Figures 2 and 4 show the logic behind our experiments. We look both at referential ambiguity when the referent is a person (who the speaker is talking to) and when the referent is an object (what the speaker is talking about). In Experiments 1–2 children watched an interaction between other agents and had to infer to whom the protagonist was talking (what does *you* refer to; Figure 2). These experiments also test whether children can reason about cost differences in third-party interactions. In Experiments 3–4 children watched a puppet request an object ambiguously, and they had to infer which object the protagonist was talking about (what does *that* refer to; Figure 4), which also tests if children can reason about cost tradeoffs in first-party interactions.

Sample Characteristics and Approach to Analyses

Due to conceptual and practical limitations of null-hypothesis significance testing (Bakan, 1966; Cohen, 2016; Cumming, 2014) we take an estimation approach to data analysis. All effect sizes include 95% CIs estimated by bootstrapping the data. We consider confidence intervals that do not cross chance performance to suggest that the underlying effect is reliably above or below chance.

All children in these experiments were recruited and tested at an urban children's museum in Boston, MA. Although most of the children were white and middle class, a range of ethnicities and socioeconomic backgrounds are represented in museum attendees overall (47% European American, 24% African American, 9% Asian, 17% Latino, 4% two or more races; 29% of museum attendees visit on days when there is free or discounted admission). All data were collected between August 2015 and January 2017, with the exception of Experiment 2b, which were collected between December 2018 and January 2019.

Experiment 1

In Experiment 1 we test if children expect requests for help to be directed to those who can help more easily. Here we manipulate costs externally: Both agents are equally competent at helping, but one of them has to travel a smaller distance to do so. Children in Experiment 1a learned about a protagonist who wanted to lift one block but could not. Children were then introduced to two additional puppets who could lift the block, but one was closer to it than the other (Figure 2). Both of these puppets, however, were equidistant from the protagonist. Children were also told that the protagonist asked one of two other agents for help lifting the block by saying "Can you help me?" Children were asked to whom the protagonist was talking. If children take agents' costs into account, they should determine that the protagonist was talking to the agent closest

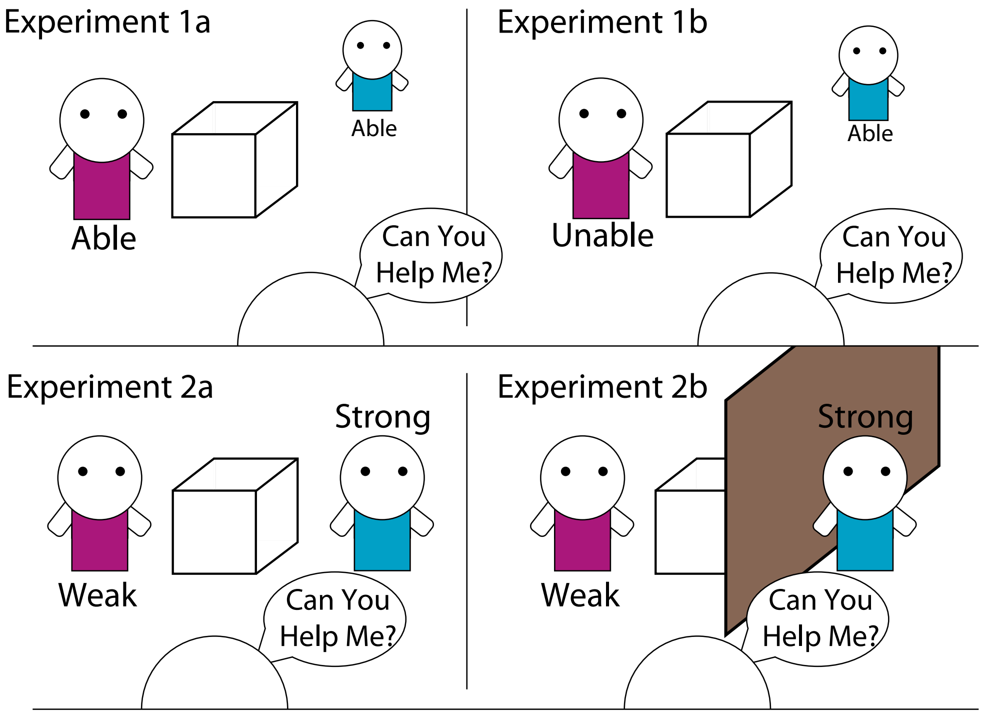

Figure 2. Simplified schematic of Experiments 1–2. In Experiment 1a children had to infer if a request for help lifting a block was directed toward an agent who was near the block or to an agent who was far away from the block. Experiment 1b was identical with the difference that the agent near the block was unable to lift it. In Experiment 2a the two agents were equidistant from the block and were both able to help, but one was stronger than the other. In Experiment 2b a wall was placed between the strong agent and the target block, such that the stronger agent was not able to help.

to the block. In contrast, if children simply expect agents to ask whoever is closer to them, they should perform at chance.

Experiment 1b used an identical setup with the difference that the puppet who was close to the block was unable to lift it (Figure 2). If children take the agent's intrinsic competence into account, they should now think that the protagonist is asking the agent who is far away from the block. However, if their judgments are purely driven by situational costs, children should continue to judge that the protagonist is talking to the agent who is close to the block.

Method

Participants

Sixteen participants (M_{age} = 5 years; 1 month, range = 4 years; 1 month–5 years; 11 months; n = 12 female) were recruited for Experiment 1a and 16 participants (M_{age} = 5 years; 0 months, range = 4 years; 3 months–5 years; 9 months; n = 7 female) were recruited for Experiment 1b. Four additional participants were recruited but not included in the study (see Coding and Exclusions).

Stimuli

The stimuli consisted of two blocks (yellow block 10.5 × 8 × 5.5 cm; green block 10.5 × 7.5 × 6.5 cm) and three puppets, which were sex matched to each participant.

Procedure

Figure 2 shows a simplified schematic of the experimental setup. Participants were tested in a quiet room in a museum. The participant and the experimenter sat on opposite sides of a small table. In Experiment 1a, a yellow block and a green block were placed on the table before the participant arrived. The experimenter introduced the protagonist puppet, and said

> Earlier today, my friend [Anne/Bob] found these two blocks—this yellow block and this green block—and s/he really wanted to build a block tower by picking up the yellow block and putting it on top of the green one, but when s/he tried to pick up the yellow block, s/he couldn't; s/he was not strong enough.

The puppet attempted and failed to lift the yellow block. The protagonist puppet was then seated on the table and the experimenter brought out two more puppets, one in each hand. The experimenter explained

> Anne/Bob's two friends arrived. When this friend tried to pick up the block, s/he could do so easily. When this friend tried to pick up the block, s/he could also pick it up easily. They were both very strong and could both lift the yellow block.

The experimenter showed the participant how the two puppets could lift the block. The experimenter then placed the two puppets such that one was sitting close to the yellow block and one was far away from the yellow block, but both puppets were equidistant from the puppet who needed help (see Figure 2). As the experimenter positioned the puppets s/he narrated:

> This friend went all the way over here. And this friend went to sit here. This friend is sitting so much closer to the block than this friend. Do you see how this friend is a lot closer to the yellow block than that friend?

This final description of the distances was included to ensure that children were aware that, although the magnitudes of the distances are negligible for humans, they were significant for the puppets. Because the room setup made it difficult to place the two puppets equidistant from the participant, we counterbalanced the location of the blocks, such that the puppet distant from the block was closer to the child on half of the trials and the puppet near the block was closer to the child on the other half of the trials. The identity of the puppet who was far away from the block, and its position relative to the child (left or right), were also counterbalanced.

Once all puppets were positioned, the experimenter said,

> remember that Anne/Bob needed help building a block tower? Well, s/he knew that both of his/her friends were strong enough to lift the block easily, but s/he also saw that one friend was a lot closer to the yellow block than the other friend. So Anne/Bob decided to ask one friend for help: "Can you help me?" Which friend did Anne/Bob ask?

The experimenter hid the protagonist right before they mentioned the ambiguous utterance to prevent children from relying on visual cues to determine the intended listener (e.g., attempting to determine who the puppet might be looking at). After children responded, they were asked an inclusion question: "Can you tell me which friend is closer to the yellow block?"

The procedure of Experiment 1b was identical to Experiment 1a with the exception that the puppet that was close to the block was unable to lift it. During the introduction, the experimenter now said,

> When this friend tried to pick up the block, s/he could do so easily. When this friend tried to pick up the block, s/he couldn't; s/he was not strong enough. So this friend is very strong and can lift the yellow block, and this friend is very weak and cannot lift the yellow block.

The experimenter demonstrated to the participant that the able puppet could lift the block and that the unable puppet failed to lift the block. The order in which the puppets were introduced (strong first or weak first) and their position relative to the participant (left or right) were counterbalanced. To match the language in Experiment 1a, the experimenter described

> This friend went all the way over here. And this friend went to sit here. This friend is sitting so much closer to the block than this friend. Do you see how this friend is a lot closer to the yellow block than that friend?

After the protagonist's friends were seated (one close to the yellow block and one far away and both equidistant from the protagonist), the experimenter said

> remember that Anne/Bob needed help building a block tower? Well, s/he knew that only this friend was strong enough to lift the block easily, but s/he also saw that this friend was a lot closer to the yellow block than the other friend. So Anne/Bob decided to ask one friend for help: "Can you help me?" Which friend did Anne/Bob ask?

As in Experiment 1a, the experimenter hid the protagonist when they asked the ambiguous test question. After children responded, they were asked an inclusion question to ensure they had followed the story: "Can you tell me which friend is stronger?"

Coding and Exclusions

After the session, the experimenter noted any script errors as well as the child's responses. Results were then coded a second time in a two-step process. For most cases, in which parents consented to videotaping (97.2% of participants), a naïve coder watched the experimental procedure and determined whether the script was run correctly, prior to viewing the participant's response to either the test or inclusion questions. The participant's responses to the test and inclusion questions were coded next. Participants were coded as not providing an answer if they failed to respond within 30 s. In cases where parents only consented to audio recording (2.8% of participants), responses were coded from audio, applying the same standards for coding. The coder and the experimenter notes had 100% agreement on all inclusion criteria and test question coding. Two participants were excluded because of an experimenter error. One participant was coded as not providing an answer because they did not respond within 30 s of the test question, and another was excluded for failing the inclusion question.

Results

Of the sixteen children included in Experiment 1a, 12 said that the request was directed to the agent who was closer to the blocks (75%; 95% CI [56.00, 100], see Figure 3). This pattern was reversed in Experiment 1b. Of the 16 participants who made a choice in Experiment 1b, 14 judged that the request for help was now directed toward the puppet that was far away but able to lift the blocks (87.5%; 95% CI [75.00, 100], Figure 3). Responses in these two experiments were reliably different from each other ($\beta = 3.39$; 95% CI [1.50, 5.71], in a logistic regression predicting puppet choice as a function of experiment, see Supporting Information). Note that in both Experiment 1a and Experiment 1b the experimenter always highlighted the distant puppet after highlighting the puppet near the block. If children's responses were driven by the order in which puppets are mentioned, they should have performed identically in both experiments, but they did not.

Experiment 2

Experiment 1 suggests that children expect requests for help to be directed toward agents who can help more easily (lower costs) and that they can use this to interpret utterances that are, strictly speaking, ambiguous. In Experiment 2 we test the same idea when the costs are intrinsic to the agents rather than determined by external factors such as distance. Children in Experiment 2a watched an incompetent puppet ask for help lifting a block. Two puppets were equidistant from the block and were both able to help, but one was stronger than the other (Figure 2). As in Experiment 1, children were asked to infer who the request for help was directed toward. We predicted that children would judge that the request was directed to the strong agent because they can help more easily.

Related research suggests that children have a baseline preference for more competent agents (e.g., Jara-Ettinger, Tenenbaum et al., 2015; Rakoczy, Hamann, Warneken, & Tomasello, 2010; Stenberg, 2013). Thus, it is possible that children assume that the protagonist is talking to the stronger agent for reasons unrelated to their ability to help in this specific situation. We test this possibility in Experiment 2b. This experiment was identical to Experiment 2a, with the exception that we introduced a wall between the strong agent and the block that the protagonist needed help lifting (Figure 2). If children expect the protagonist to talk to the stronger puppet simply because they are more competent, they should continue to select the strong agent in both experiments. But if they expect the protagonist to ask the stronger puppet because of the cost of helping, they should infer that the utterance is directed to the strong puppet in Experiment 2a and to the weak puppet in Experiment 2b.

Method

Participants

Sixteen participants (M_{age} = 5 years; 3 months, range = 4 years; 2 months–5 years; 9 months; n = 14 female) were recruited for Experiment 2a and 16 participants (M_{age} = 4 years; 11 months, range = 4 years; 2 months–5 years; 8 months; n = 7 female) were recruited for Experiment 2b. Twelve additional participants were recruited but not included in the study (see Coding and Exclusions).

Stimuli

The stimuli were the same used in Experiment 1 with the addition of a cardboard wall (42 × 42 cm) used in Experiment 2b.

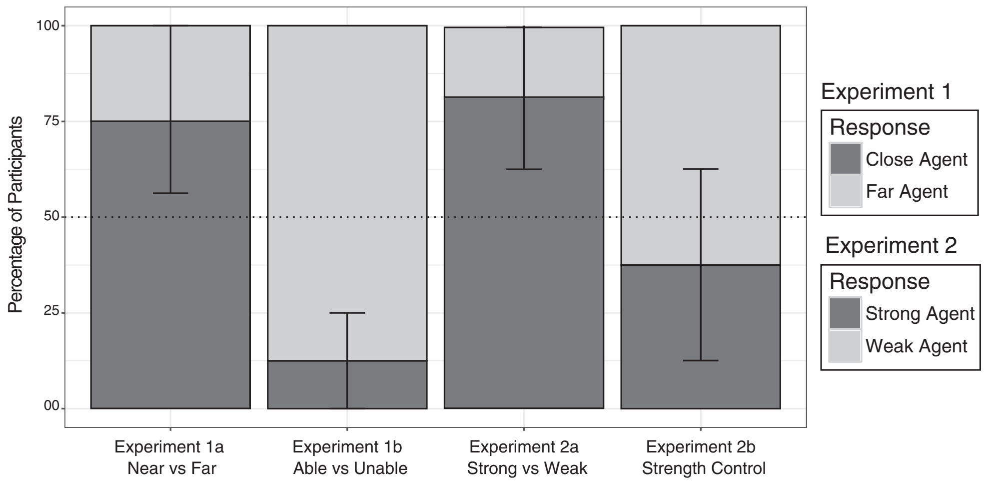

Figure 3. Results from Experiments 1–2. The *x*-axis shows each experiment and the *y*-axis shows the proportion of children judgments. Black vertical lines represent 95% CIs obtained by bootstrapping the data.

Procedure

Experiment 2a began in an identical way to Experiment 1a. After children saw that the protagonist could not lift the block, the two additional puppets were introduced. In contrast to Experiment 1a, the experimenter now showed that both puppets could lift the block, but one was stronger than the other. While the experimenter acted the situation out, s/he narrated, "When this friend tried to pick up the yellow block, s/he could do it easily. When this friend tried to pick up the yellow block, it was really hard for him/her, but s/he was able to lift it." The strong puppet was shown lifting the yellow block immediately, whereas the weak puppet struggled, made effort sounds, and eventually succeeded. Next, both puppets were placed equidistant from both the yellow block and from the protagonist. The experimenter then said,

> remember that Anne/Bob needed help building a block tower? S/he knew that this friend was stronger than this friend and could lift the block easier than that friend could. So Anne/Bob decided to ask one friend for help. "Can you help me?" Which friend did Anne/Bob ask?

As in Experiments 1a and 1b, the experimenter hid the protagonist before revealing the ambiguous utterance. After children responded, they were asked an inclusion question: "Can you tell me which friend is stronger?"

In Experiment 2b, a cardboard wall was placed between the strong agent and the target block. This experiment began in the same way as Experiment 2a. After the relative competence of the two puppets was introduced, the puppets were placed equidistant from the block and from the participant. Additionally, a cardboard wall was placed between the strong puppet and the block. The experimenter then said,

> Remember that Anne needed help building a block tower? Anne knew that both friends can lift the block, and that this friend is stronger than this friend. But she also knew that this friend is behind a big wall, so she cannot come over to lift the block. She cannot help right now. So Anne decided to ask *one* of his/her friends, for help. "Can you help me?" Which friend did Anne ask?

As in Experiment 2a, the experimenter hid the puppet before revealing the ambiguous utterance. In contrast to Experiment 2a, Experiment 2b used only female puppets rather that puppets sex-matched to the participant. After children responded, they were asked two inclusion questions: "Can you remind me, can this friend [pointing at strong agent] walk over to help now?" followed by, "Which friend is stronger?"

Coding and Exclusions

Results were coded in the same way as Experiment 1 (88.89% were available on video; 7.4% on audio). In cases where parents did not consent to audio or video (3.7% of participants), the experimenter's notes were used to determine the child's inclusion and performance. The coder and the experimenter notes had 100% agreement on all inclusion and test question coding. Three participants were excluded from the study because of an experimenter error, and nine additional participants were excluded for failing an inclusion question (with all nine of these exclusions occurring in Experiment 2b).

Results

Figure 3 shows the results from the experiment. Of 16 participants included in Experiment 2a, 13 answered that the request for help was directed toward the stronger agent (81.25%; 95% CI [62.5, 100]). The results were reversed in Experiment 2b. Of 16 children included in Experiment 2b, 10 answered that the request for help was directed to the weaker agent (62.5%; 95% CI [37.5, 87.5]). Children's responses across the two experiments were reliably different from each other ($\beta = 2.16$; 95% CI [0.59, 3.95], in a logistic regression predicting puppet choice as a function of experiment, see Supporting Information).

In Experiment 2b, out of the nine excluded participants, one failed to remember which puppet was stronger. The remaining eight excluded participants said that the strong puppet was still able to help, despite the presence of the wall, suggesting that some children did not find the wall an unsurmountable barrier. Although the results in Experiment 2b were not reliably different from chance (50% preference contained in the 95% CI [37.5, 87.5]), this may be in part to the manipulation not being sufficiently effective. Nonetheless, these results establish that children's responses in Experiment 2a cannot be explained by a low-level expectation that agents prefer to interact with more competent agents regardless of whether they are in a position to help. Note that in both Experiment 2a and Experiment 2b the experimenter always highlighted the strong puppet last. If children had relied on order effects, they should have performed identically in both experiments, but they did not.

Interim Discussion

Experiments 1 and 2 suggest that children rely on their Naïve Utility Calculus to resolve referential ambiguities when it remains unclear who the intended listener is (i.e., who does *you* refer to?). However, these experiments do not reveal whether children resolved the ambiguity, or whether they did not realize the utterance was ambiguous in the first place. The context of the event may have allowed children to predict who the protagonist would ask for help, allowing them to preemptively avoid entertaining any potential ambiguity. Under this view, when the puppet asked, "Can you help me?" children immediately assumed that the puppet was talking to the agent that they predicted would help, without realizing that the pronoun "you" was ambiguous in the linguistic context. Although this is consistent with our account because it shows that children relied on commonsense psychology to build expectations about who the protagonist would ask for help, it does not show whether children explicitly processed and resolved the ambiguity.

In Experiments 3–4 we focus on referential ambiguities where the possible referents are objects (i.e., what does *that* refer to?). In these experiments, the puppet unambiguously speaks to the participant but does not specify what he wants. In these experiments, context alone is not enough to determine what the puppet may want, making it impossible by design for children to determine what the puppet will request before he speaks. Thus, these experiments simultaneously test if (a) children can continue to resolve referential ambiguities when the appropriate inference cannot be predicted by the context alone, (b) whether children can rely on their Naïve Utility Calculus to resolve referential ambiguities when the referent is an object, and (c) if they can consider both their own costs and the costs of others in the same physical environment in order to effectively reason about the cost tradeoffs.

Experiment 3

In Experiment 3 children watched a puppet decide which of three hats to wear to a party. An orange hat was near the puppet, and an orange and a green hat were near the child (color of the hat near the puppet counterbalanced; Figure 4). The puppet asked the child to pass him "that one." If children take relative costs into account when they interpret the utterance, they should pass the hat of the color that was not near the puppet.

Method

Participants

Sixteen participants (M_{age} = 4 years; 11 months, range = 4 years; 1 month–5 years; 11 months; $n = 10$ female) were recruited and tested at an urban children's museum. Four additional

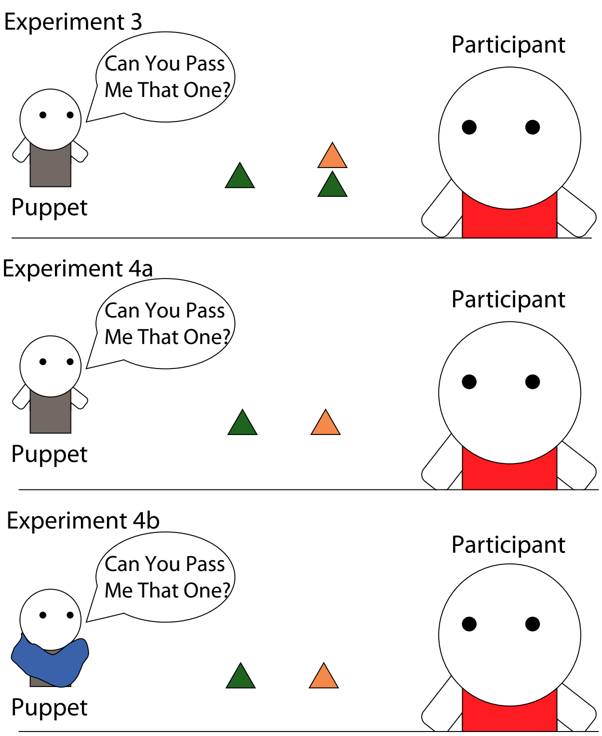

Figure 4. In Experiments 3–4 a puppet asked participants if they could pass him "that one." In Experiment 3, there was one green hat that was closer to the puppet, and a green and an orange hat closer to the participant (color counterbalanced). In Experiment 4a there was one green hat closer to the puppet and one orange hat closer to the participant (color counterbalanced). Experiment 4b was conceptually similar to Experiment 4a with the difference that the puppet now had a blanket around her arms, thus being unable to reach for either of the hats.

participants were recruited but not included in the study (see Coding and Exclusions).

Stimuli

The stimuli consisted of a puppet, two orange hats, and one green hat; or one orange hat and two green hats, depending on the condition (see Procedure).

Procedure

Children were seated across from the experimenter, on opposite sides of a small table. The experimenter had one small cardboard hat in front of him/her (orange or green, counterbalanced) and the child had a green and an orange hat placed in front of them (left/right position counterbalanced). The experimenter began by introducing a puppet (Bert from sesame street) and narrating: "Here we have some hats, and here is my friend Bert! Bert is going to a party today and he wants to wear a green hat or an orange hat, but we do not know which one." The experimenter then said, "That green/orange hat is close to Bert. That green hat and that orange hat are far away from Bert," while pointing at the hats. The experimenter then said, "Bert looked at all the hats, and he said 'Can you pass me that one?' Can you pass Bert the hat he wants?" The experimenter simultaneously moved Bert's head to indicate looking at each object and then positioned the puppet in the center looking straight ahead. The script was narrated in past tense to minimize the concern that participants would assume they could solve the task by tracking Bert's eye gaze.

Coding and Exclusions

Results were coded in the same way as Experiments 1 and 2 (100% of data available from video), with the additional constraint that the coder ensured the puppet was centered and looking straight ahead in an ambiguous manner when he requested a hat. Because Experiment 3 did not include an inclusion question, a coder blind to the child's final answer coded whether the participant was paying attention to the task. The coder and the experimenter notes had 100% agreement on test question coding. Four participants were excluded from the final sample because of experimenter error (n = 2), because the participant declined to answer the test question (n = 1), and because a coder blind to their final choice determined that the child was looking elsewhere and not paying attention to the task (n = 1).

Results and Discussion

Of 16 participants who made a choice, 13 participants took the hat near them that did not match the color of the hat near Bert (e.g., the orange hat in Figure 4) and handed it over to him (81.25%; 95% CI [62.5, 100]; Figure 5).

Experiment 4

Results from Experiment 3 suggest that children can use their Naïve Utility Calculus to infer to which object an agent is referring. It is possible, however, that children succeed in the task through a Gricean implicature (Grice, 1975). When children hear the request for "that" hat, they may infer that if the puppet wanted a hat of the common color type (e.g., the green hat in Figure 4), he would have needed to specify *which* of the two hats of that color he wanted. Thus, if the speaker follows the Gricean maxim of quantity, then he must be referring to the unique hat. Indeed, prior work has shown that 4-year-olds and adults can use this expectation to recover speakers' intended referents (Stiller, Goodman, & Frank, 2015).

In Experiment 4a we test an additional prediction that our account makes but that the simple Gricean account does not. As in Experiment 3, children watched a puppet choose which hat to wear to a party. One orange hat was near the puppet, and one green hat was near the child (colors counterbalanced across participants). The puppet asked the child to pass him "that one" (see Figure 4). If children can only solve these tasks through Gricean implicatures, they should perform at chance on this task. However, if children can also perform cost-based inferences, they should pass the hat that is close to them and far from the puppet.

Nonetheless, children may solve this task by relying on expectations about deictics alone. Even though the experimenter refers to all hats using the pronoun "that" (see Procedure), children may believe that this pronoun is more suitable for farther-away objects, and they may interpret the puppet's request accordingly (although see Reuter & Lew-Williams, 2018 for evidence suggesting this understanding emerges later). In Experiment 4b we test one final prediction of our account that controls for this possibility. Experiment 4b was identical to Experiment 4a with the difference that the puppet now wore a blanket around his arms and could therefore not reach either of the hats. If children solve these referential ambiguities by reasoning through a Naïve Utility Calculus, they should be unable to infer which hat the puppet is talking about and perform at chance. If instead children rely on interpreting "that" as referring to the faraway object, their performance should be indistinguishable from performance in Experiment 4a.

Method

Participants

Sixteen participants (M_{age} = 4 years; 10 months, range = 4 years; 0 months–5 years; 10 months; n = 5 female and n = 2 not sex coded) were recruited for Experiment 4a, and 16 participants (M_{age} = 5 years;1 months, range = 4 years;0 months–5 years; 10 months; n = 6 female and n = 1 not sex coded) were recruited for Experiment 4b. Six additional participants were recruited but not included in the study (see Coding and Exclusions).

Stimuli

The stimuli consisted of a puppet (Bert from sesame street), one orange hat and one green hat (see Procedure) and a small scarf wrapped around the puppet as a blanket in Experiment 4b.

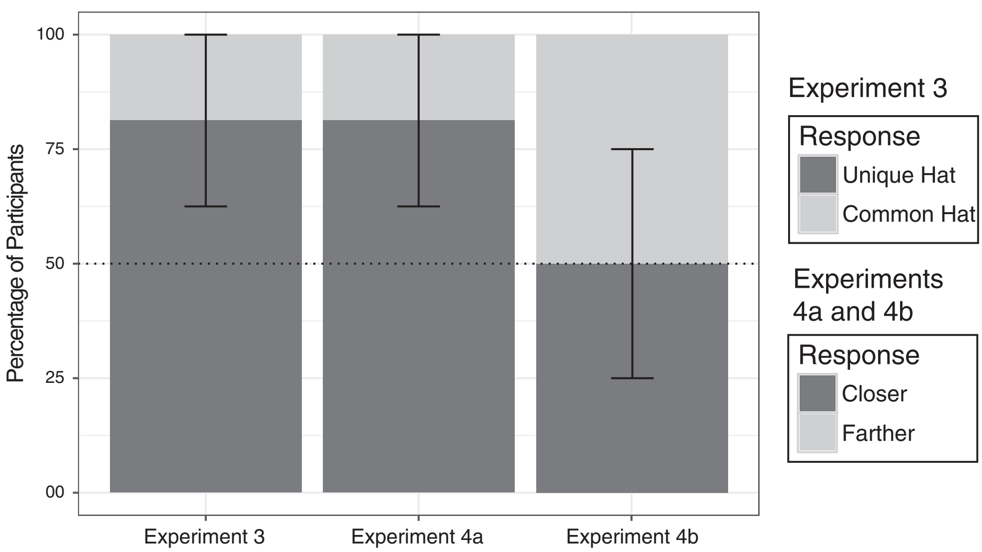

Figure 5. Results from Experiments 3–4. The *x*-axis shows each experiment and the *y*-axis shows the proportion of children judgments. Black vertical lines represent 95% CIs obtained by bootstrapping the data.

Procedure

Experiments 4a and 4b began in an identical way. Children were seated across from the experimenter, on opposite sides of a small table. The experimenter had one small cardboard hat in front of them (orange or green, counterbalanced), whereas the child had a hat placed in front of them. The hat placed close to the child was different from the color of the hat close to the experimenter (see Figure 4). The experimenter began by pulling out a puppet of Bert and saying, "Here we have some hats, and here is my friend Bert! Bert is going to a party today and he wants to wear a green hat or an orange hat, but we do not know which one." The experimenter next described, "That green/orange hat is close to Bert. That orange/green hat is far away from Bert," while sequentially pointing at the hats. In Experiment 4a the experimenter then said, "Bert looked at all the hats, and he said 'Can you pass me that one?' Can you pass Bert the hat he wants?" In Experiment 4b, the experimenter first said "but Bert has a blanket around his hands, so can't reach anything. He can't reach either of the hats." Followed by the same final prompt from Experiment 4a.

Coding and Exclusions

Results were coded in the same way as Experiments 1–3 (86.8% from video; 5.3% from audio; and 7.9% did no consent to audio or video). The coder and the experimenter had 100% agreement on test question coding. Five participants were excluded from the study because they moved the position of the hats before the puppet had made a request, and one participant was excluded because of an experimenter error.

Results and Discussion

Of 16 participants who made a choice in Experiment 4a, 13 participants passed the hat that was closest to them (81.25%; 95% CI [62.5, 100]). By contrast, out of the 16 participants who made a choice in Experiment 4b, only eight passed the hat that was closest to them (50%; 95% CI [25.00, 75.00]), showing that children were unable to infer which hat Bert was talking about. Children's responses across the two experiments were reliably different from each other ($\beta = 1.60$; 95% CI [0.02, 3.29], in a logistic regression predicting hat choice as a function of experiment, see Supporting Information; note, however, that this difference is not significant under a Fisher's exact test; $p = .14$). These results show that children's responses in Experiment 4a were not driven by expectations about deictics alone. Note that in Experiments 3 and 4 the correct hat was always mentioned last (although not immediately before the test question). The results from Experiments 1 and 2 already suggest that children did not rely on order to solve this task. If participants relied on the order in which the hats were mentioned, their performance should have been identical in Experiments 4a and 4b, but this was not the case.

As in Experiment 3, here we explicitly referred to each hat as "that hat" during the introduction. Although is possible children believe that "that" is more likely to be applied to far away items (Fillmore, 1997; Levinson, 2004; Tanz, 1980), the results from Experiment 4b suggests this was not the case in our task. If children interpret "that hat" as the more distant one, they should have passed the hat that was closest to them. Instead, children performed at chance. Note, however, that this data could be consistent with a combination of behaviors. Half of the participants may have relied on a proximity understanding of "that," whereas the other half defaulted to passing the hat that was closest to the puppet. However, research has found that four- and five-year-olds generally do not take proximity into account when interpreting deictics such as "that" and "these" (Reuter & Lew-Williams, 2018), suggesting that our participants are unlikely to have been using a cue such as proximity.

Finally, although the results from Experiment 4a show that children's inferences in our task depend on cost-based reasoning, our findings do not imply that children do not rely on Gricean implicatures at all, or in Experiment 3 specifically. Indeed, other reference resolution tasks that that can be explained through Gricean implicatures cannot be explained by the Naïve Utility Calculus (Stiller et al., 2015). Our results only suggest that children can solve ambiguities through the Naïve Utility Calculus, but this ability is not exclusive with other routes to pragmatic inference.

Analysis of Age Trends

Our overall pattern of data suggests that 4- and 5-year-olds can rely on commonsense psychology to reason about underinformative utterances. It is possible, however, that this understanding develops. We tested for this possibility by pooling data from all experiments where the Naïve Utility Calculus predicts performance different from chance (all

Experiments except 4b) and running a logistic regression with children's response as the dependent variable and their age (coded as a continuous variable) as the independent variable. We found no effect of age ($\beta = 0.79$; 95% CI [−0.15, 1.75], see Supporting Information), suggesting that children's overall success is not due only to the older participants in our sample.

Discussion

In these studies, we looked at whether 4- and 5-year-olds rely on their Naïve Utility Calculus to resolve referential ambiguities. We found that children expect ambiguous requests for help to be directed toward individuals who can help more easily, both when the costs are given by situational factors (Experiments 1a and 1b) and when they are given by the agents' competence (Experiment 2a and 2b). Children also expected agents to ask for help with things that are costlier for them to achieve relative to others in the scene (Experiments 3, 4a, and 4b). Together, children's performance on all seven experiments shows that children can resolve referential ambiguities by reasoning about cost tradeoffs between the speaker and the listener.

Our results are consistent with looking-time studies with adults showing that, when a speaker requests an object, listeners first constrain their visual search to areas that the speaker cannot reach (either because of the distance, or because she is holding objects in her hands; Hanna & Tanenhaus, 2004). Related work has also shown that 21-month-olds constrain their interpretation of referents based on what the requester can reach (Grosse, Moll, & Tomasello, 2010). In this study, an experimenter sat with a battery in front of them and a second battery across the room. The experimenter held a flashlight and asked the child to pass them "the battery" (the study used other pairs of objects in addition to the flashlight and battery). Children were more likely to hand the nearby battery when the experimenter's hands were occupied and the far-away battery when the experimenter's hands were free. These inferences can also be directly explained by the Naïve Utility Calculus and are conceptually similar to Experiments 4a and 4b. The present work goes beyond these findings in several key ways. Although many accounts of language have emphasized the role of context and cooperativeness in language (Baldwin, 1991; Bloom, 2000; Horowitz & Frank, 2016; Southgate et al., 2010), contextual information, to our knowledge, is rarely formalized or operationalized. Our framework shows how aspects of context can be represented in terms of the relative costs they impose to different agents, and this allows us to generate predictions about communicative inferences in different contexts. Similarly, cooperativeness in communication is typically explored in relation to an assumption of truthfulness (Grice, 1975). Our work expands this notion of cooperativeness to include sensitivity to the costs we impose on others. This formalization allows us to show the inferences children perform in our task and in Grosse et al (2010) can be understood as deriving from language-independent expectations about speakers' goals in different contexts, providing a novel framework for understanding the role of context and cooperation in communication more generally.

Although here we only tested qualitative predictions of the Naïve Utility Calculus, these ideas can be formalized to generate testable quantitative predictions (Rubio-Fernández, & Jara-Ettinger, 2018; Jara-Ettinger, Schulz, & Tenenbaum, under review; Jern et al., 2017). In particular, recent work has shown how a model that interprets language through the lens of a Naïve Utility Calculus can jointly resolve referential ambiguity and infer common ground in a similar way to human adults (e.g., explaining ambiguity in speakers by inferring that they may not be aware of the objects whose presence create the ambiguity; Rubio-Fernández, & Jara-Ettinger, 2018). Thus, this framework lends itself to generating quantitative predictions that can be tested developmentally and compared with adult inferences to better understand the development of reference resolution and pragmatics.

Although our study provides evidence that the Naïve Utility Calculus supports language understanding, it also faces limitations. In the first two experiments, we do not disambiguate between two explanations for children's expectation that agents should ask more competent agents for help. First, children may reason that if the agent asking for help cares about the helpers, she should prefer asking those who can help more easily. Alternatively, children may believe that agents who can help more easily are more likely to agree to help. The Naïve Utility Calculus predicts that both of these expectations should be at play. Our work did not disambiguate the contribution of each of these two expectations on children's inferences. A second limitation is that, although our experiments overtly manipulated costs, we did not manipulate relative rewards. Nonetheless, our account does make predictions about children's reasoning when reward trade-offs are at play. For instance, in the presence of a number of unlabeled holiday gifts, it would typically be infelicitous for someone to tell a conversation partner, “You can have that one.” Suppose, instead, that the speaker knows that the interlocutor specifically wanted one of the gifts. In this case, the Naïve Utility Calculus predicts that the listener can infer what the speaker means based on their mutual knowledge of the interlocutor’s rewards. Finally, our work is limited in that we exclusively focus on costs as determined by physical movement, but communicative events usually combine other types of costs, such as speech itself. For instance, it is possible that we expect agents to abbreviate or substitute shorter names in cases where a production is long, and thus costly (Mahowald, Fedorenko, Piantadosi & Gibson, 2013). As listeners, awareness of these costs may help us in reference resolution, though future work is needed to determine to what extent people apply the same reasoning with nonphysical costs.

In addition, our work touches on a distinction between varieties of ambiguity. Objectively, some utterances are underspecified and consistent with multiple referential assignments. However, only a subset of these *objectively* ambiguous utterances is also *subjectively* ambiguous, where listeners recognize this ambiguity and actively work to resolve it. Our studies focus on cases of objective ambiguity, but the extent to which participants experienced subjective ambiguity is unknown. However, our results raise the possibility that, in some situations, the Naïve Utility Calculus can prevent objective ambiguity from becoming subjective ambiguity. In our first two experiments specifically, children’s representation of the event may have helped them predict who the protagonist would ask for help, entirely circumventing the observation that expressions such as “can you help me?” were referentially ambiguous. In our last two experiments, however, it was impossible to predict which hat the protagonist would want until they requested one of the hats. Although unlikely, it is technically is possible that, as soon as the puppet began speaking, children quickly inferred what the puppet wanted before hearing the ambiguous referent. If so, this would suggest that the role of the Naïve Utility Calculus in helping avoid subjective ambiguity is even more powerful than we have proposed here.

Developmentally, recent research suggests that even infants can perform basic preference inferences predicted by the Naïve Utility Calculus (Liu et al., 2017). This, however, does not imply that the Naïve Utility Calculus does not develop or that it is readily available for language understanding. This and related work on preference inferences have manipulated costs as determined by physical effort. As noted above, however, everyday action understanding and language comprehension requires attending to other sources of costs such as attention, memory (e.g., remembering the location of objects or uncommon names of objects), and utterance length. It remains unknown how children’s understanding of costs that are not determined by physical movement develop. Moreover, successful communication in the contexts we explored requires going beyond reasoning about other agents’ costs and rewards, and it involves recognizing that these costs and rewards are in common ground. Although these representations may develop, our work provides initial evidence that 4- and 5-year-olds may track situational (e.g., relative distances to objects) and intrinsic (e.g., relative strength and subjective preferences) costs and rewards in common ground.

More broadly, our findings show that language understanding is intertwined with commonsense psychology even in early childhood. Children consider not just what other people see and know but also how people are likely to behave in different contexts based on the costs different plans impose to different agents. Their expectations of how costly or rewarding actions are likely to be in different contexts helps children interpret what people say, and this understanding may in turn, help them build more nuanced theories of other people’s behavior.

References

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 1–29. https://doi.org/10.1037/h0020412

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875–890. https://doi.org/10.2307/1131140

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/3577.001.0001

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, *12*, 104–108. https://doi.org/10.1111/cdep.12269

Bohn, M., Zimmermann, L., Call, J., & Tomasello, M. (2018). The social-cognitive basis of infants' reference to absent entities. *Cognition*, *177*, 41–48. https://doi.org/10.1016/j.cognition.2018.03.024

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 25–49). Basingstoke, UK:

Palgrave Macmillan. https://doi.org/10.1057/9780230524125_2

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*(1), 19–41. https://doi.org/10.1080/01690968708406350

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*, 245–258. https://doi.org/10.1016/S0022-5371(83)90189-5

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Cohen, J. (2016). The earth is round (p<.05). In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 69–82). New York, NY: Routledge.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Fillmore, C. J. (1997). *Lectures on deixis*. Stanford, CA: CSLI.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in Cognitive Sciences*, *7*, 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Graf, E., & Davies, C. (2014). The production and comprehension of referring expressions. *Pragmatic Development in First Language Acquisition*, *10*, 161. https://doi.org/10.1075/tilar.10.10gra

Grice, H. P. (1975). *Logic and conversation*. New York, NY: Academic Press. https://doi.org/10.25291/VR/1975-VR-58

Grosse, G., Moll, H., & Tomasello, M. (2010). 21-month-olds understand the cooperative logic of requests. *Journal of Pragmatics*, *42*, 3377–3383. https://doi.org/10.1016/j.pragma.2010.05.005

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105–115. https://doi.org/10.1207/s15516709cog2801_5

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*, 807–819. https://doi.org/10.1111/cdev.12527

Jara-Ettinger, J., Floyd, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children believe that agents maximize expected utilities. *Journal of Experimental Psychology: General*, *146*, 1574. https://doi.org/10.1037/xge0000345

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*, 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14–23. https://doi.org/10.1016/j.cognition.2015.03.006

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (under review). *The naïve utility calculus as a unified, quantitative framework for action understanding*.

Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science*, *26*, 633–640. https://doi.org/10.1177/0956797615572806

Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition*, *168*, 46–64. https://doi.org/10.1016/j.cognition.2017.06.017

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*, 179–208. https://doi.org/10.1016/S0010-0277(03)00002-7

Levinson, S. (2004). Deixis. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 97–121). Oxford, UK: Blackwell. https://doi.org/10.1002/9780470756959.ch5

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, *15*, 545–556. https://doi.org/10.1111/j.1532-7078.2009.00028.x

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*, 1038–1041. https://doi.org/10.1126/science.aag2132

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., . . . Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PLoS ONE*, *9*, e92160. https://doi.org/10.1371/journal.pone.0092160

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*, 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*, 826–835. https://doi.org/10.1111/j.1467-7687.2007.00615.x

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*, 603–613. https://doi.org/10.1348/026151005X55370

Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and over-informative utterances? *Journal of Pragmatics*, *59*, 26–39. https://doi.org/10.1016/j.pragma.2013.03.007

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*, 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–677. https://doi.org/10.2307/1131839

Pesowski, M. L., Denison, S., & Friedman, O. (2016). Young children infer preferences from a single action, but not if it is constrained. *Cognition*, *155*, 168–175. https://doi.org/10.1016/j.cognition.2016.07.004

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*, 785–798. https://doi.org/10.1348/026151009X479178

Reuter, T., & Lew-Williams, C. (2018). *Look at THAT: Using deixis to evaluate prediction-via-simulation with L1 children, L1 adults, and L2 adults*. Talk presented at BUCLD 43, Boston University, Boston, MA.

Rubio-Fernández, P., & Jara-Ettinger, J. (2018). Joint inferences of speakers' beliefs and referents based on how they speak. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, *43*, 696–704. https://doi.org/10.1037/0012-1649.43.3.696

Saylor, M. M., Ganea, P. A., & Vázquez, M. D. (2011). What's mine is mine: Twelve-month-olds use possessive pronouns to identify referents. *Developmental Science*, *14*, 859–864. https://doi.org/10.1111/j.1467-7687.2010.01034.x

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*, 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Stenberg, G. (2013). Do 12-month-old infants trust a competent adult? *Infancy*, *18*, 873–904. https://doi.org/10.1111/infa.12011

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190. https://doi.org/10.1080/15475441.2014.927328

Tanz, C. (1980). *Studies in the acquisition of deictic terms*. Cambridge, UK: Cambridge University Press.

Vaish, A., Demir, Ö. E., & Baldwin, D. (2011). Thirteen- and 18-month-old infants recognize when they need referential information. *Social Development*, *20*, 431–449. https://doi.org/10.1111/j.1467-9507.2010.00601.x

Wilson, D., & Sperber, D. (2002). Relevance theory. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford, UK: Blackwell.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Analysis of Differences Across Conditions

Appendix S2. Analysis of Age Effects

Appendix S3. Experiment Scripts