Cognition 250 (2024) 105836

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Children use disagreement to infer what happened☆

Jamie Amemiya [a,*], Gail D. Heyman [b], Tobias Gerstenberg [c]

[a] Department of Psychology, Occidental College, USA
[b] Department of Psychology, University of California, San Diego, USA
[c] Department of Psychology, Stanford University, USA

ARTICLE INFO

Keywords:
Disagreement
Inference
Prediction
Theory of mind
Ambiguous speech

ABSTRACT

In a rapidly changing and diverse world, the ability to reason about conflicting perspectives is critical for effective communication, collaboration, and critical thinking. The current pre-registered experiments with children ages 7 to 11 years investigated the developmental foundations of this ability through a novel social reasoning paradigm and a computational approach. In the inference task, children were asked to figure out what happened based on whether two speakers agreed or disagreed in their interpretation. In the prediction task, children were provided information about what happened and asked to predict whether two speakers will agree or disagree. Together, these experiments assessed children's understanding that disagreement often results from ambiguity about what happened, and that ambiguity about what happened is often predictive of disagreement. Experiment 1 ($N = 52$) showed that children are more likely to infer that an ambiguous utterance occurred after learning that people disagreed (versus agreed) about what happened and found that these inferences become stronger with age. Experiment 2 ($N = 110$) similarly found age-related change in children's inferences and also showed that children could reason in the forward direction, predicting that an ambiguous utterance would lead to disagreement. A computational model indicated that although children's ability to predict when disagreements might arise may be critical for making the reverse inferences, it did not fully account for age-related change.

1. Introduction

Solving the most fundamental problems of our time requires that people work together productively, find common ground, and negotiate solutions. In order to do this successfully, they must be able to understand and make inferences about conflicting perspectives. Here, we examine the developmental origins of this ability by assessing children's inferences about the kinds of disagreements that are likely to occur in everyday social life.

Imagine, for example, that you overhear the following disagreement at the park: One person is sure that Sam wanted Robin to paint the wagon, while the other person is sure that Sam did *not* want Robin to paint the wagon. Given that you have no reason to weigh one person's perspective more than the other, you may infer that Sam must have said something *ambiguous* that resulted in different interpretations. Maybe Sam said something like, “My wagon would look better in a new color,” which could have been interpreted as either an indirect request for Robin to paint it or merely an observation (Ackerman, 1978).

The ability to consider ambiguous events as causes of disagreement is critical for navigating the complexity of social life. Indeed, a less optimal alternative would be to always privilege one perspective (e.g., only considering the first perspective that Sam wanted the wagon to be painted), which could lead to inaccurate inferences about what happened. In two pre-registered experiments, we examined children's ability to infer ambiguous utterances from disagreement across ages 7 to 11 years old. In addition, we examined the mechanisms underlying this inference. Specifically, we applied a Bayesian model to test whether children's ability to *predict* that ambiguous utterances will cause disagreement underlies their inferential reasoning.

1.1. Inferences from disagreement

Prior research on disagreement has primarily focused on how children decide which of two conflicting perspectives is more likely to be right (Harris et al., 2018; Koenig et al., 2019; Poulin-Dubois & Brosseau-Liard, 2016). In this work, children are given cues that indicate one person is more trustworthy and their perspective should be privileged. Studies find that children successfully make use of a range of cues,

☆ All the preregistrations, data, study materials, and analysis code are available here: https://github.com/cicl-stanford/children_disagree

* Corresponding author at: Occidental College, Department of Psychology, 1600 Campus Rd, Los Angeles, CA 90041, USA
E-mail address: amemiya@oxy.edu (J. Amemiya).

https://doi.org/10.1016/j.cognition.2024.105836
Received 13 November 2023; Received in revised form 9 April 2024; Accepted 23 May 2024

including each informant's history of providing accurate information (Bazhydai et al., 2020; Ronfard et al., 2018) and relevant domain expertise (Boseovski & Thurman, 2014; Lane & Harris, 2015). Children also recognize that personal biases may be the cause of certain disagreements, such as when people disagree about who should win a contest (Mills et al., 2012; Mills & Grant, 2009; Mills & Keil, 2008).

Less research has examined how children resolve disagreements in which there is no clear basis for privileging one perspective over another. Thus, there are open questions as to whether children can use disagreement to infer that an ambiguous event occurred. Several recent studies inform this question.

One study examined if children ages 4 to 9 can suspend judgment when they and another speaker have equally strong evidence for different claims (Langenhoff et al., 2023). The authors found it is not until age 7 to 9 that children explicitly recognize they should suspend judgment and seek more information. A related study examined whether children ages 4 to 9 recognize that both speakers can be right in certain disagreements (Foushee & Srinivasan, 2017). For example, children observed speakers disagree about whether an object was "tall" or "not tall," which could be accounted for by the fact that the speakers had observed different object distributions (i.e., one speaker had only observed short versions of the object, while the other had only observed tall versions). Similar to Langenhoff et al. (2023), it was not until around age 8 to 9 that children started to acknowledge that both speakers could be right.

To our knowledge, only one study has directly examined children's ability to infer ambiguity from disagreement. Amemiya et al. (2021) asked 5- to 12-year-old children to infer which object was being discussed after hearing a disagreement about its description. For example, children heard one person call a hidden object "pink," and the other person call it "orange." Children were then presented with a pink object, an orange object, and an object that was somewhat pink and orange and thus more ambiguous with respect to its color label. The authors found that children did not choose the ambiguous object above chance until around age 9 to 10, aligning with prior studies that it is not until later in childhood that children recognize disagreement as a cue to ambiguity (Foushee & Srinivasan, 2017; Langenhoff et al., 2023).

1.2. Disagreement in the social domain: the case of ambiguous speech

Studies thus far suggest that children's tendency to integrate disagreeing perspectives develops slowly across childhood. One shared feature of these prior studies is that they focus on disagreements about the properties of objects (Amemiya et al., 2021; Foushee & Srinivasan, 2017; Langenhoff et al., 2023). In the current research, we examine disagreement in the *social* domain for two reasons. First, examining a novel domain can provide insight into the generalizability of the object property work, such as whether there is also protracted development in children's ability to infer ambiguity from social disagreements. Second, understanding the complexities of disagreement is an important social skill that can help children successfully navigate their interpersonal relationships. Indeed, social actions are often ambiguous and can elicit different interpretations (Sperber & Wilson, 1987). One such context that frequently elicits different interpretations, and that is the focus of the present research, is an *indirect speech act* that may be interpreted by others as an indirect request or as a mere observation (e.g., Sam's statement, "My wagon would look better in a new color," in the opening example) (Ackerman, 1978).

Children's understanding of ambiguous utterances. Children's appreciation of ambiguity in communication emerges early in life and becomes increasingly sophisticated with age (Nilsen & Graham, 2012). By age 4, children recognize ambiguous utterances, showing greater hesitancy to act following an ambiguous versus unambiguous statement (Plumert, 1996). Studies using other measures, such as children's visual search for more information following ambiguous utterances, indicate an even earlier understanding around age 2 (Nurmsoo & Bloom, 2008). Children's understanding of ambiguous communication also becomes more robust across childhood. For example, although young children show an understanding of ambiguity when they are the direct listener, they struggle to appreciate ambiguity from a third-person perspective until around age 6 (Nilsen & Graham, 2012).

Regarding indirect speech acts, there has been much research on how children *respond* to such statements (Aguert & Laval, 2013; Carrell, 1981; Elrod, 1987; Marocchini et al., 2022). This research suggests that children's responses to indirect speech acts are context-dependent (Ackerman, 1978; Carrell, 1981; Elrod, 1987), which could indicate that children recognize their ambiguity. Specifically, when it is clarified that a person is about to make a request, children as young as 4 years of age respond appropriately to indirect statements such as, "I would love to see the circle colored blue" or "I'll be very sad if you make the circle red" (i.e., children in turn color the circle blue) (Carrell, 1981). However, without such clarification, children are less certain about the speaker's intention (Elrod, 1987). Relatedly, Ackerman (1978) found that 8-year-olds (and to a lesser extent, 6-year-olds) were able to flexibly interpret statements such as, "The garbage is beginning to smell" as either an indirect request (to take out the trash) or as a mere observation (of the smell) depending on contextual cues (e.g., if the garbage is still in the kitchen or is already outside). Taken together, children appear to recognize that indirect speech acts are ambiguous when they are in the role of the listener.

1.3. Applying a computational framework to children's inferences

Less is known about whether children use disagreement to *infer* that someone made an ambiguous utterance and what cognitive processes are involved in making this inference. To inform our hypotheses, we draw from computational work on inferential reasoning (Gerstenberg et al., 2021; Houlihan et al., 2023; Jara-Ettinger et al., 2016; Outa et al., 2022), which has broadly characterized how reasoners make inferences about unobserved causes, including past physical events and others' beliefs, goals, and desires. According to this framework, drawing accurate inferences from disagreement relies on the ability to *predict*, in the forward direction, that ambiguous utterances can cause disagreement. Based on this generative model, children can then reason about the inverse direction and *infer* that an ambiguous utterance occurred when hearing others' disagreement. Put simply, children's *predictions* underlie their *inferences*.

When considering potential age-related change, the computational framework indicates several explanations for why older children may be more successful in inferring ambiguity. First, older children may be better able to *predict* that ambiguous statements can cause disagreement. In line with this possibility, older children ages 7 to 8 years old are able to predict that ambiguous stimuli, such as the duck/rabbit illusion, can cause different interpretations among people (Beck, McColgan, et al., 2011; Carpendale & Chandler, 1996). Although children younger than 7 are able to switch between interpretations of ambiguous stimuli themselves (Gopnik & Rosati, 2001), they struggle to accept that two people can disagree and have different interpretations at the same time (Beck, Robinson, et al., 2011).

Another possibility is that young children *can* make the expected predictions, but may struggle with later steps in the inferential reasoning process. Specifically, even if children can apply their predictions to draw correct inferences about the probability of each utterance within a given situation, these inferences still need to be translated into appropriate decisions (i.e., *choosing* the most probable utterance). Doing so imposes demands on working memory and executive functions that may continue to develop well into later childhood (Best & Miller, 2010; Garon et al., 2008). In line with this possibility, computational work examining children's exploratory behavior finds that, across age 4 to 9 years old, children's actions become less noisy and more directed (Giron et al., 2023; Meder et al., 2021).

1.4. The present research

We examined children's developing ability to infer ambiguous utterances from disagreement. Given that children may not have the necessary skills to make this inference until age 7 to 8 (Beck, Robinson, et al., 2011; Carpendale & Chandler, 1996; Langenhoff et al., 2023) or potentially as late as age 10 (Amemiya et al., 2021), we included children ages 7 to 11 years old. Focusing on a slightly older age range also allowed us to better examine the mechanisms that underlie how children *successfully* infer ambiguity from disagreement.

Fig. 1 shows the conceptual model of predictive and inferential reasoning that guides our two experiments. Experiment 1 tested whether children would be more likely to infer that a person made an ambiguous utterance (vs. an unambiguous utterance) after hearing others disagree about what happened compared to when they agree. Experiment 2 sought to replicate the results of Experiment 1, as well as test a mechanism of how children make the inference. We examined children's ability to predict that ambiguous utterances will cause disagreement, while unambiguous utterances will cause agreement. We then applied a Bayesian model to examine if children's predictions explain performance on the inference task, and whether this model can help to explain potential age-related change.

2. Transparency and openness statement

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards (Kazak, 2018). All data, analysis code, and research materials are available here. We used R version 4.2.3 (R Core Team, 2023) and the following R packages: bookdown v. 0.33 (Xie, 2016, 2023a), broom.mixed v. 0.2.9.4 (Bolker & Robinson, 2022), car v. 3.1.2 (Fox & Weisberg, 2019), kableExtra v. 1.3.4 (Zhu, 2021), knitr v. 1.42 (Xie, 2014, 2015, 2023b), lme4 v. 1.1.32 (Bates et al., 2015), Metrics v. 0.1.4 (Hamner & Frasco, 2018), rmarkdown v. 2.21 (Allaire et al., 2023; Xie et al., 2018; Xie et al., 2020), rsample v. 1.1.1 (Frick et al., 2022), scales v. 1.2.1 (Wickham & Seidel, 2022), tidyverse v. 2.0.0 (Wickham et al., 2019), xtable v. 1.8.4 (Dahl et al., 2019). Both experiments were preregistered via the Open Science Framework (Experiment 1, Experiment 2).

3. Experiment 1: inference

In this experiment, we investigated children's ability to infer what a speaker had said in a social scenario, based on whether the two listeners agreed or disagreed about what happened.

3.1. Methods

Participants. Fifty-two U.S. children aged 7 to 11 years (at least 10 children in each age group; 27 girls, 25 boys; 14 Asian, 14 White, 7 Mixed, 4 Black, 4 Latinx, 3 Middle Eastern, 1 reported "Other", 5 not reported) were recruited from parks and social media. As noted in our pre-registration, we planned a sample size of 50, with at least 10 children per age group, based on a power analysis (2 more participants were run to be inclusive of children at the public park). An additional five children were dropped from analyses due to excessive noise and distractions at the park (2), parent interference (1), and parents reporting that the child is a new English speaker (2).

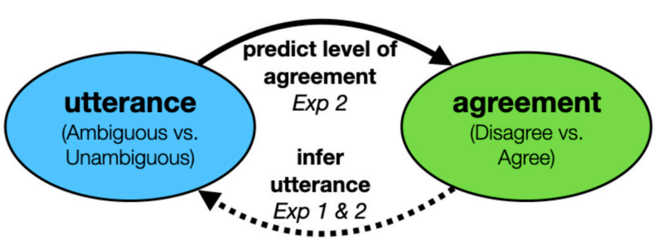

Fig. 1. Conceptual model linking utterances and agreement. Experiment 1 examined children's inferred utterances (Ambiguous vs. Unambiguous) from level of agreement (Disagree vs. Agree). Experiment 2 also examined inferences, as well as children's predictions of the level of agreement following each utterance type. The solid arrow represents a causal link, and the dashed arrow an inferential link.

Procedure. Children participated in the study one-on-one with an experimenter over Zoom (28) or at a park (24). Children first completed a practice trial without any feedback that familiarized them with the story structure.

Following the practice trial, children were presented with four test stories (wagon, tower, snowman, and dog). Fig. 2 shows an example version of the study; all of the scenarios can be found in the full Appendix here and a sample video of the procedure can be found here. In all stories, the target child uttered an unknown statement to a second child, and the second child subsequently intervened. Two adults in the story overheard the statement and, critically, either *agreed* or *disagreed* about whether the target child wanted the intervention. On agreement trials, adults both stated they were "really sure" that the child wanted the intervention. In the disagreement trials, one adult was "really sure" that the child wanted the intervention, whereas the second adult was "really sure" the child did *not* want the intervention. Agreement was manipulated within subjects, such that two of the stories were agreement trials (e.g., wagon and snowman) and two were disagreement trials (e.g., tower and dog).

At the end of each story, participants were asked to infer what the target child said and were given the following response options:

(a) an unambiguous request (e.g., "Please paint my wagon orange")
(b) an ambiguous statement (e.g., "My wagon would look better in a new color")
(c) a random statement (e.g., "My wagon has four wheels")

In the agreement trials, we were interested in whether children would infer the unambiguous request, as both informants agreed about what the child wanted. If children instead chose the ambiguous utterance, this may indicate they misinterpreted that statement as a clear request. Moreover, if they chose the random statement, this could indicate children were confused by the scenario or task in general.

In the disagreement trials, we were interested in whether children would infer the ambiguous utterance. If so, this would suggest that children are resolving the disagreement by inferring an utterance that explains both perspectives. If, on the other hand, children chose the unambiguous statement, this would indicate that they privileged one person's view (e.g., the person who was sure that Sam wanted the intervention) and disregarded the other. Finally, if children chose the random statement in the disagreement trials, this could mean that they avoided choosing the unambiguous statement, but were unsure about which of the two remaining statements explain the disagreement.

We counterbalanced several factors to reduce the possibility of other design factors influencing the results (8 versions in total):

(1) Stories were presented in one of two orders (either wagon to dog or dog to wagon).
(2) Trial types were presented in one of two orders (either [agreement, disagreement, agreement, disagreement] or [disagreement, agreement, disagreement, agreement]). In this way, stories were crossed with trial type (e.g., the version in Fig. 2 had the wagon story as the agreement trial, while another version had the wagon story as the disagreement trial).
(3) We varied the *valence* of the unambiguous request and subsequent agreement remarks (either the valence was *positive* as described in the example version, or *negative*, such that the

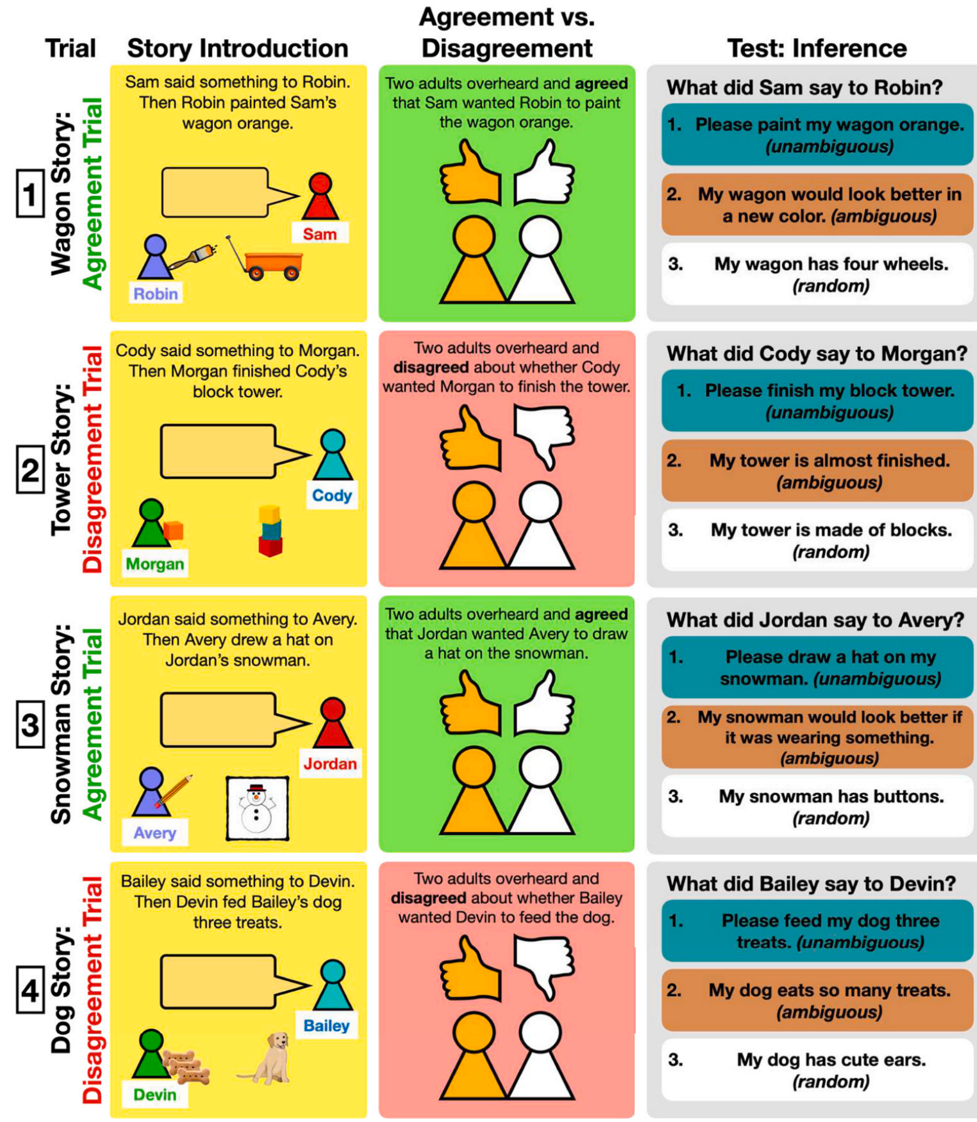

Fig. 2. Experiment 1: Example study version (there were 8 versions in total). Each story had two adult observers either agree or disagree in their interpretation of an utterance. Children were then asked to infer what was said. Note that stories were fully crossed with Agreement/Disagreement trial type (e.g., another study version had the Wagon story as a Disagreement trial).

speaker said, “Please do *not* paint my wagon” and the listeners agreed that the speaker did *not* want the intervention).

We examined if any of these counterbalanced factors moderated the key effect of trial type. Analyses that tested an interaction between each factor and trial type indicated that none of these factors made a difference for the reported results (all 95% CIs included 0), story order*trial type: $B = -0.23$ (95% CI -0.97, 0.51); trial order*trial type: $B = -0.65$ (95% CI -1.41, 0.11); valence*trial type: $B = 0.35$ (95% CI -0.40, 1.10), thus we will not discuss these factors further.

3.2. Hypotheses

As noted in our pre-registration, we predicted an effect of trial type (agreement vs. disagreement), such that children would be more likely to infer unambiguous statements in the agreement trials (relative to the disagreement trials), and more likely to infer ambiguous utterances in the disagreement trials (relative to the agreement trials).

In a secondary set of hypotheses, we predicted that children would infer unambiguous utterances above chance (33%) in the agreement trials and infer ambiguous utterances above chance (33%) in the disagreement trials.

We also pre-registered exploratory analyses examining the moderating effect of age. Specifically, we were interested in whether children's sensitivity to the trial type (agreement vs. disagreement) may strengthen with age. We report *p*-values for our confirmatory analyses, but refrain from doing so for the exploratory analyses (following the suggestion by Wagenmakers et al., 2012).

3.3. Results

Fig. 3 presents the percentage that children inferred the ambiguous utterances by continuous age in years and trial type. Notably, we found that children almost always inferred either the unambiguous or ambiguous statements; only one child chose the random statement on one trial. Fig. 3 can thus be essentially interpreted as inferring

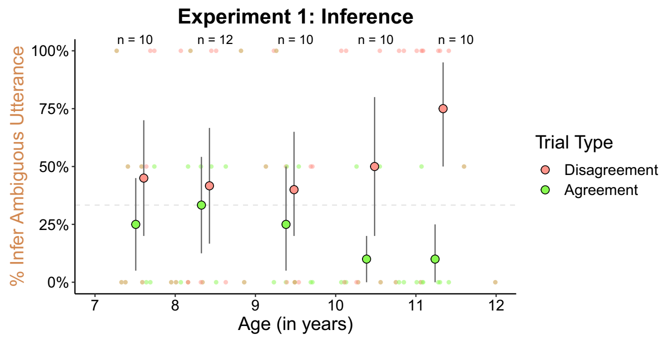

Fig. 3. Experiment 1: Percentage that participants inferred ambiguous utterances in the Disagreement and Agreement trials by continuous age in years. *Note*: Large points show mean percentages for each age group centered at the mean age for that group. Error bars are 95% bootstrapped confidence intervals. Small points show percentages for individual participants (each participant contributes two data points on this plot, one for each trial type). n = number of participants in each age group.

ambiguous vs. unambiguous utterances. This also suggested that children were choosing statements that were relevant to at least one of the testimonies they heard. Moreover, this meant that some of our preregistered analyses were redundant (i.e., predicting ambiguous statements versus other statements and predicting unambiguous statements versus other statements were simply inverses of one another). So, we only report one set of those analyses: inferring ambiguous statements.

Confirmatory analyses. We ran a mixed-effects logistic regression model predicting children's selection of the ambiguous statement (e.g., "My wagon would look better in a new color" vs. the other two statements) with a fixed effect of trial type (disagreement vs. agreement), as well as a random intercept for each participant. This model indicated that children were more likely to infer ambiguous statements after hearing disagreement compared to agreement, $B = 1.83$ (95% CI 1.03, 2.63), $p < .001$, $OR = 6.22$.

We also found that children inferred unambiguous statements above chance (33%) in the agreement trials, $\chi^2(1) = 40.14$, $p < .001$, and inferred ambiguous statements above chance (33%) in the disagreement trials, $\chi^2(1) = 4.87$, $p = .027$.

Exploratory analyses. To explore potential differences by age, we ran a model that included the main effect of trial type, main effect of age (continuous), and the interaction between trial type and age. We found that there was a sizeable interaction, $B = 0.85$ (95% CI 0.26, 1.43), $OR = 2.33$, indicating that the disagreement effect strengthened with age.

We next explored at what age children distinguished between agreement and disagreement trials by running the mixed-effects model for each age group. This analysis indicated that the trial type effect was found only among the 10-year-old age group, $B = 3.23$ (95% CI 0.57, 5.88), and the 11-year-old age group, $B = 4.53$ (95% CI 1.18, 7.87). We did not find trial type effects in age groups younger than 10 (all 95% CIs included 0).

3.4. Discussion

Experiment 1 found that children's ability to infer ambiguous utterances from disagreement is slow-developing, even in the social domain: It was not until 10 years of age that children distinguished between disagreement versus agreement trials. This finding aligns with prior work documenting that children fail to infer ambiguous *objects* from disagreement until about age 10 (Amemiya et al., 2021), and that children's ability to reason about ambiguity and multiple perspectives continues to develop into later childhood (Beck, Robinson, et al., 2011; Carpendale & Chandler, 1996; Langenhoff et al., 2023).

With respect to younger children, the data indicated that they sometimes selected the ambiguous statements, but on other trials, they chose the unambiguous statement that aligned with only one of the informants' perspectives. This result is consistent with prior work finding that younger children often have trouble integrating two conflicting views (Amemiya et al., 2021; Langenhoff et al., 2023). Experiment 2 sought to replicate these findings and to test a mechanism for how children successfully infer ambiguous statements.

4. Experiment 2: prediction and inference

Informed by computational and developmental work on inferential reasoning (Gerstenberg et al., 2021; Houlihan et al., 2023; Jara-Ettinger et al., 2016; Outa et al., 2022), Experiment 2 examined how children's ability to predict that ambiguous utterances can lead to disagreement may explain their performance on the inference task.

Children were randomly assigned to either the prediction or inference condition. The prediction condition presented children with utterances that were either unambiguous (e.g., "Please paint my wagon") or ambiguous (e.g., "My wagon would look better in a new color"). They were then asked if two listeners would agree or disagree in their interpretation of the statement. Fig. 4 shows an example pair of prediction trials in the wagon story. Following unambiguous statements, we were interested in whether children predict that the listeners will agree rather than disagree. Following ambiguous statements, we were interested in whether children would be *more* likely to predict that the listeners will disagree compared to the unambiguous statement trials. However, we expected that children's *overall* rate of predicting disagreement following ambiguous statements would be around 50%. This is because each listener is essentially a coin flip in their interpretation of an ambiguous statement, resulting in a 50% chance that they will agree and a 50% chance that they will disagree.

Regarding the inference condition, we made some changes from Experiment 1. First, we removed the random option given that participants rarely chose it. Second, we had participants reason about *both* types of trials for each story for our proposed analyses (see below for more on the Bayesian model). For example, participants now reasoned about an agreement wagon story *and* a disagreement wagon story. Across both the prediction and inference conditions, participants completed eight test trials in total (four pairs of stories).

We explored to what extent a Bayesian model accurately characterizes how children make inferences on this task across age. Specifically, for each age group, we linked predictions to inferences using Bayes' theorem:

$$p(utterance|agree) \propto p(agree|utterance)p(utterance) \tag{1}$$

We assumed that the prior on utterances p(utterance) was uniform (unambiguous and ambiguous utterances are equally likely). The model uses children's data from the prediction condition for the likelihood p(agree|utterance) to compute the posterior p(utterance|agree), which represents what responses in the inference condition should look like if children were reasoning in line with Bayesian inference.

Let us illustrate the model's predictions via a concrete example. Consider a situation in which there are two different possible utterances, an unambiguous statement or an ambiguous statement. We assume that the probability of agreement given an unambiguous statement is high with p(agree|utterance = unambiguous) = 0.8. In contrast, if a statement is ambiguous, we expect that two people are just as likely to agree with one another about its meaning as they are to disagree, such that p(agree|utterance = ambiguous) = 0.5. Assuming that, a priori, an ambiguous utterance is just as likely as an unambiguous one, we can now compute the probability that an utterance was ambiguous under the two possible scenarios. When two people agreed with one another, the

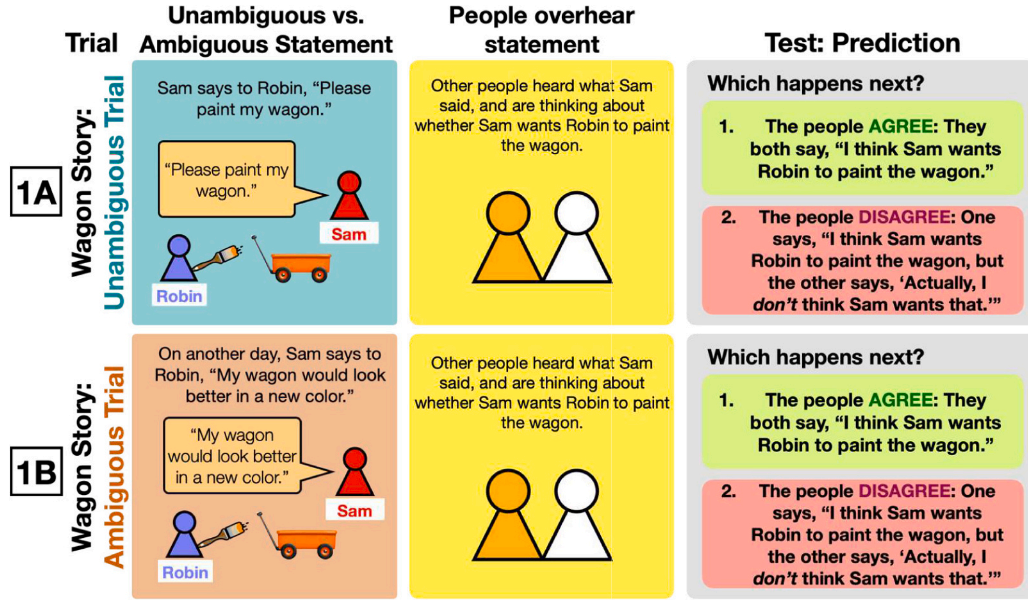

Fig. 4. Experiment 2: Sample pair of prediction test stories (participants completed four pairs in total). Each story presented either an unambiguous or ambiguous utterance. Children were then asked to predict whether two listeners would agree or disagree.

probability that the utterance was ambiguous is:

$$p(ambiguous|agree) = \frac{p(agree|ambiguous) \cdot p(ambiguous)}{p(agree|ambiguous) \cdot p(ambiguous) + p(agree|unambiguous) \cdot p(unambiguous)} = \frac{p(0.5) \cdot p(0.5)}{p(0.5) \cdot p(0.5) + p(0.8) \cdot p(0.5)} \approx 0.38$$

In contrast, if two people disagreed with one another, the probability that the utterance was ambiguous is:

$$p(ambiguous|disagree) = \frac{p(0.5) \cdot p(0.5)}{p(0.5) \cdot p(0.5) + p(0.2) \cdot p(0.5)} \approx 0.71$$

In our experiment, participants do not provide a continuous response but instead choose a binary label saying whether the statement was ambiguous or unambiguous. Hence, we need to translate from the inferred continuous belief to a discrete choice. Different versions of the model either assume that this choice is directly proportional to the posterior probability of the two options, or that the choice is determined via a softmax decision function that's based on the inferred beliefs:

$$p_{choose}(ambiguous|agree) = \frac{e^{\beta \cdot p_{belief}(ambiguous|agree)}}{\sum_{i \in agreement} e^{\beta \cdot p_{belief}(ambiguous|agree)}} \quad (2)$$

The temperature parameter β in the softmax decision function determines how likely the model chooses one of the two options based on its posterior beliefs. For example, if the β parameter is very high, then the model deterministically chooses one option even if it is only slightly more likely than the alternative. So even if an ambiguous statement was only slightly more likely than an unambiguous statement, it would still always choose the ambiguous statement. In contrast, if the β parameter was 0, then the model would randomly choose between the two options. For other values of β, the model will choose probabilistically based on the inferred beliefs – it will be more likely to choose the option with the greater posterior belief, but there is still some chance that it will choose the other option instead. For our example, if two listeners agreed, the probability that the model will choose the ambiguous statement p_{choose}(ambiguous|agree) would be 27% (assuming that p_{belief}(ambiguous|agree) = 0.38, p_{belief}(ambiguous|disagree) = 0.71, and $\beta = 3$).

Below, we will compare several versions of the model against participants' responses in our experiment. The different versions of the model all compute their inferences according to Bayes' rule (see Eq. 1) but differ in how these inferences inform their choices (Eq. 2).

4.1. Methods

Participants. Participants were 110 U.S. children aged 7 to 11 years (53 girls, 55 boys, 1 non-binary, 1 not reported; 42 Asian, 31 White, 17 Mixed, 14 Latinx, 3 Black, 1 Middle Eastern, 2 not reported) recruited from public parks, social media, and MIT's Lookit platform for live studies (Scott & Schulz, 2017). Approximately half of the participants (56) were randomly assigned to the prediction condition, while the remaining half were assigned to the inference condition (54). As noted in our pre-registration, we planned for a sample size of 100, with at least 10 children per age group, and had slightly more children due to over-recruitment on Lookit. An additional four children were dropped due to failing attention checks (2), significant difficulties with attention (1), or the parent reporting that the child is a new English speaker (1).

Procedure. Children completed the study one-on-one with a live experimenter over Zoom (104) or at a public park (6). After completing a practice trial, children were presented with four stories (i.e., wagon, tower, snowman, and dog). Each story had two trials (e.g., the wagon story had two trials), resulting in eight trials total. In the prediction condition, children reasoned about an unambiguous statement and an ambiguous statement for each story. In the inference condition, children reasoned about an agreement and a disagreement for each story. All of the scenarios can be found in the full Appendix here, and a sample video of the prediction condition can be found here and the inference condition here.

At the end of each prediction trial, children were asked to predict whether the listeners would (a) agree, or (b) disagree in their interpretations of the utterance. For the inference trials, children were asked to infer what the child said: (a) an unambiguous request (e.g., "Please paint my wagon"), or (b) an ambiguous statement (e.g., "My wagon would look better in a new color").

Similar to Experiment 1, we counterbalanced the following factors:

(1) The order of the stories (either from wagon to dog or from dog to wagon).
(2) The order of whether the unambiguous statement/agreement versus the ambiguous statement/disagreement trial came first.
(3) The valence of the unambiguous request and subsequent agreement remarks (either positive, e.g., "Please paint my wagon" or negative, e.g., "Please do *not* paint my wagon").

There were 16 study versions in total from counterbalancing these factors (there were twice as many versions because we now had an inference *and* a prediction condition). These factors did not moderate the effect of trial type for either the prediction condition, story order*trial type: $B = 0.26$ (95% CI -0.17, 0.69); trial order*trial type: $B = 0.02$ (95% CI -0.41, 0.44); valence*trial type: $B = 0.34$ (95% CI -0.08, 0.77), or the inference condition, story order*trial type: $B = -0.26$ (95% CI -0.86), 0.33; trial order*trial type: $B = 0.08$ (95% CI -0.52, 0.68); valence*trial type: $B = -0.31$ (95% CI -0.91, 0.30), and thus we will not discuss these factors further.

4.2. Hypotheses

As noted in our pre-registration, we predicted an effect of trial type in both the prediction and inference conditions. Specifically, for the prediction condition, we hypothesized that children would be more likely to predict disagreement following ambiguous statements (vs. unambiguous statements). For the inference condition, we hypothesized that, as in Experiment 1, children would be more likely to infer ambiguous utterances in the disagreement trials (relative to the agreement trials).

We also report the following exploratory analyses:

(1) The moderating effect of age.
(2) Whether Bayesian inference characterizes children's reasoning.

4.3. Results

Fig. 5 shows children's prediction data, while Fig. 6 shows children's inference data. The points overlaying the data in Fig. 6 represent the model's predictions for children's inferences using the data from the prediction condition.

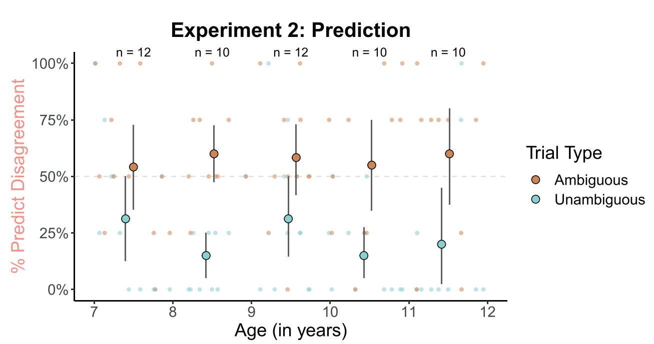

Fig. 5. Experiment 2: Percentage that participants predicted disagreement in the Ambiguous Utterance and Unambiguous Utterance trials by continuous age in years. *Note*: Large points show mean percentages for each age group centered at the mean age for that group. Error bars are 95% bootstrapped confidence intervals. Small points show percentages for individual participants (each participant contributes two data points on this plot, one for each trial type). n = number of participants in each age group.

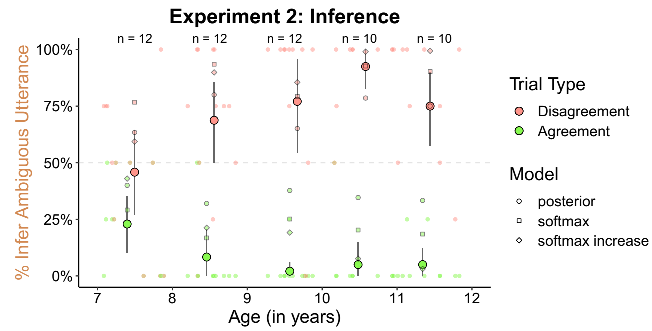

Fig. 6. Experiment 2: Percentage that participants inferred ambiguous utterances in the Disagreement and Agreement trials by continuous age in years. *Note*: Large points show mean percentages for each age group centered at the mean age for that group. Error bars are 95% bootstrapped confidence intervals. Symbols with black outline show model predictions based on participants' predictions as shown in Fig. 5. Small points show percentages for individual participants (each participant contributes two data points on this plot, one for each trial type). n = number of participants in each age group.

Confirmatory analyses. For the prediction condition, we ran a mixed-effects logistic regression model that examined children's prediction of disagreement (vs. agreement), with a fixed effect of trial type (ambiguous vs. unambiguous statement) and a random intercept for each participant. As hypothesized, children were more likely to predict disagreement in the ambiguous statement trials relative to the unambiguous statement trials, $B = 1.58$ (95% CI 1.14, 2.02), $p < .001$, $OR = 4.87$. Furthermore, the rates of predicting each outcome aligned with expectations: In the unambiguous statement trials, children predicted agreement 77% of the time (i.e., close to 100%); while in the ambiguous statement trials, children predicted agreement 43% of the time and disagreement 57% of the time (i.e., close to 50% for each outcome).

For the inference condition, we ran a mixed-effects logistic regression model that examined children's inference of an ambiguous statement (vs. unambiguous), with a fixed effect of trial type (disagree vs. agree) and a random intercept for each participant. As expected, children were more likely to infer an ambiguous statement in the disagreement trials, $B = 3.77$ (95% CI 3.08, 4.46), $p < .001$, $OR = 43.42$. This result replicated the findings in Experiment 1 that also found a main effect of trial type on children's inferences.

Exploratory analyses: Moderation by age. To test for moderation effects by age, we ran additional mixed-effect models that included the main effect of trial type, the main effect of age, and the interaction term between trial type and age (continuous) for the prediction and inference conditions. We then explored at which age children successfully distinguished between trial types by running the mixed-effects model within each age group.

For the prediction condition, we found that there was no moderating effect of age, $B = 0.21$ (95% CI −0.09, 0.51), $OR = 1.23$, indicating that the effect of trial type (ambiguous vs. unambiguous statement) was consistent across age. In line with the interaction indicating no age moderation effect, even 7-year-olds distinguished between unambiguous and ambiguous statements when making predictions, $B = 1.21$ (95% CI 0.25, 2.17), $OR = 3.35$.

For the inference condition, we found that there was a moderating effect of age, $B = 1.27$ (95% CI 0.74, 1.80), $OR = 3.57$, indicating that the effect of trial type (disagree vs. agree) on inferring ambiguous statements strengthened with age. We explored when children began to distinguish between disagreement and agreement and found that even 7-year-olds made this distinction, $B = 1.21$ (95% CI 0.24, 2.16), $OR = 3.31$. Thus, while we replicated Experiment 1's result that the effect of trial type strengthens with age, we found that the *emergence* of this effect was earlier in Experiment 2 (7 years old) than in Experiment 1 (10 years old). We return to potential explanations in the discussion.

Exploratory analyses: Bayesian inference. Fig. 6 shows model predictions that link from children's predictions to inferences in three different ways. The "posterior" model simply applies Bayes' rule as stated in Eq. 1. This model doesn't have any free parameters as it directly uses the judgments from participants in the prediction condition. The model captures the main trends in the data: children are most likely to infer unambiguous statements in agreement trials, and ambiguous statements in disagreement trials. However, this version of the model overestimates children's inferences of ambiguous utterances for agreement trials.

We fit two additional models that are inspired by the Rational Speech Acts framework (Degen, 2023; Goodman & Frank, 2016). Here, speakers and listeners are expected to make decisions about what utterance to use, or what situation to infer from an utterance, based on their joint goal of communicating successfully. In these models, it's a standard assumption that speakers (and listeners) are softmax-rational (Luce, 1959; Sutton & Barto, 1998). Speakers choose utterances based on their expected utility. The softmax decision function has a temperature parameter (β) which captures how likely a speaker will communicate the expression with the greater expected utility.

We fit a "softmax" model which assumes that children use their posterior inferences to make a decision about which statement to choose at test (see Eq. 2). This model has one free parameter: the temperature parameter in the softmax function. We fit this parameter by minimizing the sum of squared error between model predictions and averaged inferences. Introducing a softmax transformation shifts down the predictions for agreement trials and thus brings them closer in line with the data. It also shifts up the predictions for the disagreement trials. However, the model still slightly overestimates ambiguous inferences for agreement trials.

The "softmax increase" model assumes that the temperature parameter in the softmax function changes as a linear function of age. This model has two free parameters: one for the intercept and one for the slope in the linear function that maps from age to the temperature parameter β. This model predicts that older children are more likely to infer the option with the higher posterior compared to younger children. Although the model now tends to overpredict childrens' inferences of the ambiguous statement for the disagreement trials, the model arguably captures children's inferences for the agreement trials best.

In terms of model fit, the "posterior" model has a correlation of $r = 0.96$ and a root mean squared error (RMSE) of 0.21. By comparison, the values for the "softmax" model are $r = 0.96$, RMSE $= 0.17$, and for the "softmax increase" model they are $r = 0.98$, RMSE $= 0.15$.[1]

4.4. Discussion

Experiment 2 replicated the two main results from Experiment 1: Children were more likely to infer ambiguous statements after hearing disagreement versus agreement, and this ability strengthened with age. Interestingly, however, we found an earlier emergence of this ability, such that children as young as 7 were able to make this inference (Experiment 1 found that it was not until children were age 10). One explanation is that providing fewer options (we removed the third, random statement) and observing *both* types of story trials (e.g., children heard both the agreement and disagreement versions of the wagon story) helped children focus on the contrast between the unambiguous and ambiguous statements. Even with these additional scaffolds, however, it is notable that we still found age-related improvements in children's inferential reasoning.

We next turn to the prediction condition, which we included to understand how children make inferences and the pattern of age-related change. First, we found support for the hypothesis that children successfully predict that ambiguous utterances are more likely than unambiguous utterances to cause disagreement. This ability emerged at 7 years old. However, in contrast to the inference condition, we did not find age effects; 7-year-olds performed just as well as the oldest children. Across age groups, children's predictions followed the expected pattern in which they favored agreement after the unambiguous statements (about 75% of the time) whereas they were roughly split in their predictions following ambiguous statements. The consistency in this pattern across age groups provided evidence against the account that age-related change in children's prediction abilities explains the age patterns in inferences. However, it still left open the possibility that there may be age-related differences later in the inferential process.

To examine how children's predictions relate to their inferences more directly, we applied several computational models within a Bayesian framework. The "posterior" model, which simply applies Bayes' rule, captured the main effect of trial type (disagreement vs. agreement). This suggests that, in line with prior computational work in other domains of cognition, children's predictions play an important role in their inferences (Gerstenberg et al., 2021; Houlihan et al., 2023; Jara-Ettinger et al., 2016; Outa et al., 2022). However, this model did *not* capture age-related change in inferences, due to the fact that children's predictions did not follow the same age-related increases. Instead, we found that the "softmax increase" model better captured the age-related changes in inferences. This points to the explanation that older children are more likely to choose the utterance that aligns with their posterior beliefs.

5. General discussion

Children develop in a world where disagreement is commonplace. Prior work has mostly examined *which speaker* children believe in a disagreement. Here, we shift the focus to inferences about *what happened*, investigating whether children use disagreement to infer that someone uttered an ambiguous statement. We find evidence across two experiments that children's ability to make this inference strengthens across 7 to 11 years of age. Moreover, our computational results indicate that children's ability to *predict* that ambiguous utterances can cause disagreement underlies their inferences and that age-related improvement in inferential reasoning may be due to being able to choose the utterance aligning with their posterior beliefs.

Our research builds upon the previous finding that children's ability to infer ambiguity from disagreements about *objects* develops slowly across childhood (Amemiya et al., 2021). We find a similar prolonged trajectory when children reason about *social* disagreements, suggesting that children still find making this inference challenging despite likely having a lot of experience with disagreement in this domain. The consistent developmental pattern across object and social scenarios suggests a domain-general age-related increase in the ability to use disagreement as a cue to ambiguity. In this way, our work aligns with previous studies suggesting that children's understanding of ambiguity becomes more robust across late childhood (Carpendale & Chandler, 1996; Foushee & Srinivasan, 2017; Langenhoff et al., 2023).

The current research also sheds new light on *how* children successfully infer ambiguity from disagreement. In particular, the simplest "posterior" model applying Bayes' rule indicates that children's predictions in the forward direction can account for the key pattern in their inferences (i.e., inferring ambiguity more after hearing disagreement than agreement). These findings support prior computational work indicating that Bayesian inference broadly characterizes children's inferential reasoning across many domains (Gerstenberg et al., 2021; Houlihan et al., 2023; Jara-Ettinger et al., 2016; Outa et al., 2022). Moreover, this result suggests a potential strategy to improve children's inferential reasoning–specifically, this inference may be boosted by strengthening children's ability to *predict* that ambiguous stimuli can cause disagreement.

Our findings also inform why older children are more successful in making this inference. First, we find that this pattern does *not* appear to be driven by age-related increases in making the correct predictions,

[1] Because of the small number of data points (10 in total), we didn't perform any statistical comparisons between the models.

which is stable across 7 to 11 years old. This stable trend aligns with previous work finding that, by around age 7, children appreciate that ambiguous stimuli (e.g., the duck/rabbit illusion) can cause people to have different interpretations (Carpendale & Chandler, 1996). Second, it also does not appear that younger children in our sample simply default to one perspective when hearing disagreement (instead, they were closer to chance). Rather, we find evidence from the "softmax increase" model that older children are more successful in choosing the right utterance based on their posterior beliefs. Similar results have been found in studies of children's exploratory behavior (Giron et al., 2023; Meder et al., 2021), in which older children more reliably choose optimal actions based on their beliefs. We posit that age-related change in making optimal choices may be due to improvements in working memory and other executive functions (Best & Miller, 2010; Garon et al., 2008). It will be important to examine whether age-related change earlier in life (e.g., from 4 to 7 years old) may be driven by different mechanisms, in light of research indicating that there are changes in children's prediction abilities (Beck, Robinson, et al., 2011) and their tendency to default to one person's perspective (Amemiya et al., 2021; Foushee & Srinivasan, 2017; Langenhoff et al., 2023) during this period.

To the extent that younger children's difficulties in inferential reasoning may be due to a noisier action selection process, constraining the number of alternatives may facilitate better performance. Notably, we find some support for this possibility when comparing the age of emergence across Experiments 1 and 2. Specifically, we found an *earlier* age of emergence in Experiment 2, in which children were presented with fewer options in the inference task (i.e., the random option was removed). Although children rarely chose the random option, it is possible that the inclusion of this distractor made it more challenging for younger children to weigh the other two options effectively. Other research has similarly found that constraining alternatives improves young children's inferential abilities, such as their ability to evaluate underinformative pedagogy (Gweon & Asaba, 2018) and understand scalar implicatures (Skordos & Papafragou, 2016). Furthermore, children in Experiment 2 observed *both* the disagreement and agreement trials for the same exact story. By eliminating superficial differences between trials, the key contrast of disagreement versus agreement may have been much more salient. This interpretation aligns with prior research indicating that greater alignment between two cases helps children notice the relevant abstract structure (Christie & Gentner, 2010). Finally, we note that Experiment 2 described two "people" as the speakers rather than two "adults" as in Experiment 1. This raises broader questions about whether what children know about the people who disagree matters, such as whether they might reason differently about disagreement between two adults versus two children versus one child and one adult.

Because the ambiguous statements used in the present study have the potential to serve as indirect speech acts, our work informs our understanding of the development of pragmatic inference. Our findings suggest that recognizing how linguistic ambiguity can contribute to diverging mental representations is a late-developing cognitive capacity. Even after this basic capacity is in place, it is likely there are contexts in which children and adults still fail to consider ambiguity as a cause of disagreement, such as when their prior beliefs align with one of the speakers. More broadly, our findings contribute to our understanding of how children can acquire information that they were never directly taught based on their observations of what others say or do (Horowitz & Frank, 2016; Ma et al., 2023).

In conclusion, this research provides an experimental paradigm to test the developmental origins of how we make sense of a common form of disagreement–those in which both people's perspectives may be worth considering in order to figure out what happened. The study results offer new insights into the developmental trajectory and underlying mechanisms of reasoning about ambiguity and disagreement.

Funding

JA was supported by grants from NICHD F32 HD098777 and NSF 2203810. TG was supported by a grant from the Stanford Institute for Human-Centered Artificial Intelligence (HAI).

CRediT authorship contribution statement

Jamie Amemiya: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Funding acquisition. **Gail D. Heyman:** Conceptualization, Resources, Writing – review & editing, Methodology. **Tobias Gerstenberg:** Conceptualization, Formal analysis, Methodology, Resources, Software, Visualization, Writing – review & editing, Funding acquisition.

Data availability

Our data, analysis, and materials are available to the public online (https://github.com/cicl-stanford/children_disagree).

Appendix A. Appendix

Experiment 1 - Version 1: Positive Valence, DADA, Wagon to Dog

Training story. Let's play a detective game! In this game, you are trying to fill in the missing part of a story. Let's try one!

You'll meet two kids. For this story, the kids are Casey and Terry. Casey said something, and you'll try to figure out what Casey said based on the clues that follow.

So, next in the story, Terry pushed Casey down the slide. Two adults saw the whole thing happen: They heard what Casey said and they saw Terry push Casey down the slide. The adults were asked: Do you think that Casey wanted to be pushed down the slide, based on what Casey said? One adult said, "Yes, I'm really sure that Casey wanted to be pushed down the slide!" The other adult agreed and said, "Yes, I'm really sure that Casey wanted to be pushed down the slide!"

Check 1: Can you remind me, what did the first adult say? [child responds] So the first adult said, "Yes, I'm really sure Casey wanted to be pushed down the slide."

Check 2: Can you remind me, what did the second adult say? [child responds] So the second adult said, "Yes, I'm really sure Casey wanted to be pushed down the slide."

Test: So based on what the adults said, let's try to figure out what Casey said right before Terry pushed Casey down the slide. Remember, the adults agreed with each other that they were really sure that Casey wanted to be pushed down the slide. Which of these three things did Casey say right before Terry pushed Casey down the slide?

1. "Please push me down the slide." (direct request)
2. "I go down slides sometimes." (ambiguous utterance)
3. "The slide is green." (random statement)

(No feedback was provided.) That was practice. We will play the official guessing game four times altogether. Let's do the first one!

Story 1: Wagon, Disagreement. Sam and Robin were at the park. Sam said something. Then Robin painted Sam's wagon orange. Let's try to figure out what Sam said!

Two adults saw the whole thing happen: They heard what Sam said and they saw Robin paint Sam's wagon orange. The adults were asked: Do you think that Sam wanted Robin to paint the wagon orange, based on what Sam said?

One adult said, "Yes, I'm really sure that Sam wanted Robin to paint the wagon orange." The other adult disagreed and said, "No, I'm really sure that Sam did *not* want Robin to paint the wagon orange."

Check 1: Can you remind me, what did the first adult say? [child responds] So the first adult said, “Yes, I'm really sure that Sam wanted Robin to paint the wagon orange.”

Check 2: Can you remind me, what did the second adult say? [child responds] So the second adult said, “No, I'm really sure that Sam did *not* want Robin to paint the wagon orange.”

Test: So based on what the adults said, let's try to figure out what Sam said right before Robin painted Sam's wagon orange. Remember, the adults disagreed with each other: One adult was really sure that Sam wanted Robin to paint the wagon orange, while the other adult was really sure that Sam did *not* want Robin to paint the wagon orange.

Which of these three things did Sam say right before Robin painted Sam's wagon orange?

1. “Please paint my wagon orange.” (direct request)
2. “My wagon would look better in a new color.” (ambiguous utterance)
3. “My wagon has four wheels.” (random statement)

Story 2: Tower, Agreement. Cody and Morgan were in the play room. Cody said something. Then Morgan finished Cody's block tower. Let's try to figure out what Cody said!

Two adults saw the whole thing happen: They heard what Cody said and they saw Morgan finish Cody's block tower. The adults were asked: Do you think that Cody wanted Morgan to finish the block tower, based on what Cody said?

One adult said, “Yes, I'm really sure that Cody wanted Morgan to finish the block tower.” The other adult agreed and said, “Yes, I'm really sure that Cody wanted Morgan to finish the block tower.”

Check 1: Can you remind me, what did the first adult say? [child responds] So the first adult said, “Yes, I'm really sure that Cody wanted Morgan to finish the block tower!”

Check 2: Can you remind me, what did the second adult say? [child responds] So the second adult said, “Yes, I'm really sure that Cody wanted Morgan to finish the block tower!”

Test: So based on what the adults said, let's try to figure out what Cody said right before Morgan finished Cody's block tower. Remember, the adults agreed with each other that they were really sure that Cody wanted Morgan to finish the block tower. Which of these three things did Cody say right before Morgan finished Cody's block tower?

1. “Please finish my block tower.” (direct request)
2. “My block tower is almost finished.” (ambiguous utterance)
3. “My tower is made of blocks.” (random statement)

Story 3: Snowman, Disagreement. Jordan and Avery were in the art room. Jordan said something. Then Avery drew a hat on Jordan's snowman. Let's try to figure out what Jordan said!

Two adults saw the whole thing happen: They heard what Jordan said and they saw Avery draw a hat on Jordan's snowman. The adults were asked: Do you think that Jordan wanted Avery to draw a hat on the snowman, based on what Jordan said?

One adult said, “Yes, I'm really sure that Jordan wanted Avery to draw a hat on the snowman.” The other adult disagreed and said, “No, I'm really sure that Jordan did *not* want Avery to draw a hat on the snowman.”

Check 1: Can you remind me, what did the first adult say? [child responds] So the first adult said, “Yes, I'm really sure that Jordan wanted Avery to draw the hat on the snowman!”

Check 2: Can you remind me, what did the second adult say? [child responds] So the second adult said, “No, I'm really sure that Jordan did *not* want Avery to draw a hat on the snowman!”

Test: So based on what the adults said, let's try to figure out what Jordan said right before Avery drew a hat on Jordan's snowman. Remember, the adults disagreed with each other: One adult was really sure that Jordan wanted Avery to draw a hat on the snowman, while the other adult was really sure that Jordan did *not* want Avery to draw a hat on the snowman. Which of these three things did Jordan say right before Avery drew a hat on Jordan's snowman?

1. “Please draw a hat on my snowman.” (direct request)
2. “My snowman would look better if it was wearing something.” (ambiguous utterance)
3. “My snowman has buttons.” (random statement)

Story 4: Dog, Agreement. Bailey and Devin were at the dog park. Bailey said something. Then Devin fed Bailey's dog three treats. Let's try to figure out what Bailey said!

Two adults saw the whole thing happen: They heard what Bailey said and they saw Devin feed Bailey's dog three treats. The adults were asked: Do you think that Bailey wanted Devin to feed the dog three treats, based on what Bailey said?

One adult said, “Yes, I'm really sure that Bailey wanted Devin to feed the dog three treats.” The other adult agreed and said, “Yes, I'm really sure that Bailey wanted Devin to feed the dog three treats.”

Check 1: Can you remind me, what did the first adult say? [child responds] So the first adult said, “Yes, I'm really sure that Bailey wanted Devin to feed the dog three treats!”

Check 2: Can you remind me, what did the second adult say? [child responds] So the second adult said, “Yes, I'm really sure that Bailey wanted Devin to feed the dog three treats!”

Test: So based on what the adults said, let's try to figure out what Bailey said right before Devin feed Bailey's dog three treats. Remember, the adults agreed with each other that they were really sure that Bailey wanted Devin to the dog three treats. Which of these three things did Bailey say right before Devin feed Bailey's dog three treats?

1. “Please feed my dog three treats.” (direct request)
2. “My dog eats so many treats.” (ambiguous utterance)
3. “My dog has cute ears.” (random statement)

Experiment 2 - Prediction Version 1: Positive Valence, UAUA, Wagon to Dog

Training Story. Let's play a guessing game! In this game, you will guess what happens next in a story. Let's try one!

If I say “Cats are the best animal”, and this person *also* thinks cats are the best animal, are they going to: 1. agree or 2. disagree with what I just said?

Test: So will we: 1. agree or 2. disagree? [child responds] Yeah, we will agree!

Now if I say “Dogs are the best animal” and this person does *not* think dogs are the best animal, are they going to 1. agree or 2. disagree with what I just said?

Test: So will we 1. agree or 2. disagree? [child responds] Yeah, we will disagree!

That was practice! We will play the official guessing game four times all together. Let's do the first one!

Story 1: Wagon, U/A. Sam and Robin are at the park. Sam says to Robin,

“Please paint my wagon.”

Check 1: Can you remind me, what did Sam say to Robin? [child responds] Yeah, Sam said to Robin, “Please paint my wagon.”

Other people heard what Sam said, and are thinking about whether Sam wants Robin to paint the wagon. What happens next?

1. The people agree: They both say, “I think Sam wants Robin to paint the wagon.”
2. The people disagree: One says, “I think Sam wants Robin to paint the wagon, butthe other says, ‘Actually, I *don't* think Sam wants that.”’

Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds]

On a different day, Sam and Robin are at the park. Sam says to Robin, “My wagon would look better in a new color.”

Check 2: Can you remind me, what did Sam say to Robin? [child responds] Yeah, Sam said to Robin, "My wagon would look better in a new color."

Other people heard what Sam said, and are thinking about whether Sam wants Robin to paint the wagon. What happens next?

1. The people agree: They both say, "I think Sam wants Robin to paint the wagon."
2. The people disagree: One says, "I think Sam wants Robin to paint the wagon, butthe other says, 'Actually, I *don't* think Sam wants that."

   Test: So which happens next: Will they 1. agree, or 2. disagree? [child responds]

Story 2: Block tower, A/U. Cody and Morgan are in the playroom. Cody says to Morgan, "My block tower is almost finished."

Check 1: Can you remind me, what did Cody say to Morgan? [child responds] Yeah, Cody said to Morgan, "My block tower is almost finished."

Other people heard what Cody said, and are thinking about whether Cody wants Morgan to finish the block tower. What happens next?

1. The people agree: They both say, "I think Cody wants Morgan to finish the block

tower."

2. The people disagree: One says, "I think Cody wants Morgan to finish the blocktower, but the other says, 'Actually, I *don't* think Cody wants that."'

   Test: So which happens next: Will they 1. agree, or 2. disagree? [child responds].

On a different day, Cody and Morgan are in the playroom. Cody says to Morgan, "Please finish my block tower."

Check 2: Can you remind me, what did Cody say to Morgan? [child responds] Yeah, Cody said to Morgan, "Please finish my block tower."

Other people heard what Cody said, and are thinking about whether Cody wants Morgan to finish the block tower. What happens next?

1. The people agree: They both say, "I think Cody wants Morgan to finish the block

tower."

2. The people disagree: One says, "I think Cody wants Morgan to finish the blocktower, but the other says, 'Actually, I *don't* think Cody wants that."'

   Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds].

Story 3: Snowman, U/A. Avery and Jordan are in the art room. Avery says to Jordan, "Please draw on my snowman."

Check 1: Can you remind me, what did Avery say to Jordan? [child responds] Yeah, Avery said to Jordan, "Please draw on my snowman."

Other people heard what Avery said, and are thinking about whether Avery wants Jordan to draw on the snowman. What happens next?

1. The people agree: They both say, "I think Avery wants Jordan to draw on thesnowman."
2. The people disagree: One says, "I think Avery wants Jordan to draw on the snowman, but the other says, 'Actually, I *don't* think Avery wants that."'

Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds].

On a different day, Avery and Jordan are in the art room. Avery says to Jordan, "My snowman would look better if it was wearing something."

Check 2: Can you remind me, what did Avery say to Jordan? Yeah, Avery said to Jordan, "My snowman would look better if it was wearing something."

Other people heard what Avery said, and are thinking about whether Avery wants Jordan to draw on the snowman. What happens next?

1. The people agree: They both say, "I think Avery wants Jordan to draw on thesnowman."
2. The people disagree: One says, "I think Avery wants Jordan to draw on the snowman, but the other says, 'Actually, I *don't* think Avery wants that."'

Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds].

Story 4: Dog, A/U. Bailey and Devin are at the dog park. Bailey says to Devin, "My dog eats so many treats."

Check 1: Can you remind me, what did Bailey say to Devin? [child responds] Yeah, Bailey said to Devin, "My dog eats so many treats."

Other people heard what Bailey said, and are thinking about whether Bailey wants Devin to feed the dog some treats. What happens next?

1. The people agree: They both say, "I think Bailey wants Devin to feed the dog some treats."
2. The people disagree: One says, "I think Bailey wants Devin to feed the dog sometreats, but the other says, 'Actually, I *don't* think Bailey wants that."'

   Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds].

On a different day, Bailey and Devin are at the dog park. Bailey says to Devin, "Please feed my dog some treats."

Check 2: Can you remind me, what did Bailey say to Devin? [child responds] Yeah, Bailey said to Devin, "Please feed my dog some treats."

Other people heard what Bailey said, and are thinking about whether Bailey wants Devin to feed the dog some treats. What happens next?

1. The people agree: They both say, "I think Bailey wants Devin to feed the dog some treats."
2. The people disagree: One says, "I think Bailey wants Devin to feed the dog sometreats, but the other says, 'Actually, I *don't* think Bailey wants that."'

   Test: So, which happens next: Will they 1. agree, or 2. disagree? [child responds].

Experiment 2 - Inference Version 1: Positive Valence, ADAD, Wagon to Dog

Training Story. Let's play a guessing game! In this game you will guess what happened in a story. Let's try one!

This person said something about animals. Then, I agreed, and said, "Yes, cats are the best animal!"

Test: What did this person say: 1. "Cats are the best animal" or 2. "Dogs are the best animal"? [child responds] Yeah, the person said "Cats are the best animal."

On a different day, this person said something about animals. Then, I disagreed, and said, "Actually, dogs are *not* the best animal!"

Test: What did this person say: 1. "Cats are the best animal" or 2. "Dogs are the best animal"? [child responds] Yeah, the person said "Dogs are the best animal."

That was practice! We will play the official guessing game four times all together. Let's do the first one!

Story 1: Wagon, A/D. Sam and Robin were at the park. Sam said something to Robin about the wagon.

Other people heard what Sam said, and thought about whether Sam wants Robin to paint the wagon. The people agreed: They both said, "I think Sam *wants* Robin to paint the wagon."

Check 1: Can you remind me, did they agree or disagree about what Sam wants? [child responds] Yeah, they agreed: They both said, "I think Sam *wants* Robin to paint the wagon."

Based on this, what did Sam say to Robin?

1. "Please paint my wagon." (direct request)
2. "My wagon would look better in a new color." (ambiguous utterance)

Test: So, which did Sam say to Robin: 1 or 2? [child responds].

On a different day, Sam and Robin were at the park. Sam said something to Robin about the wagon.

Other people heard what Sam said, and thought about whether Sam wants Robin to paint the wagon. The people disagreed: One said, "I think Sam *wants* Robin to paint the wagon, but the other said, "Actually, I *don't* think Sam wants that."

Check 2: Can you remind me, did they agree or disagree about what Sam wants? [child responds] Yeah, they disagreed: One said, "I think Sam *wants* Robin to paint the wagon, but the other said, 'Actually, I *don't* think Sam wants that."' Based on this, what did Sam say to Robin?

1. "Please paint my wagon." (direct request)
2. "My wagon would look better in a new color." (ambiguous utterance)

Test: So, which did Sam say to Robin: 1 or 2? [child responds]

Story 2: Block tower, D/A. Cody and Morgan were in the play room. Cody said something to Morgan about the block tower.

Other people heard what Cody said, and thought about whether Cody wants Morgan to finish the block tower. The people disagreed: One said, "I think Cody *wants* Morgan to finish the block tower, but the other said, 'Actually, I *don't* think Cody wants that."'

Check 1: Can you remind me, did they agree or disagree about what Cody wants? [child responds] Yeah, they disagreed: One said, "I think Cody *wants* Morgan to finish the block tower, but the other said, 'Actually, I *don't* think Cody wants that."' Based on this, what did Cody say to Morgan?

1. "Please finish my block tower." (direct request)
2. "My block tower is almost finished." (ambiguous utterance)

Test: So, which did Cody say to Morgan: 1 or 2? [child responds].

On a different day, Cody and Morgan were in the playroom. Cody said something to Morgan about the block tower.

Other people heard what Cody said, and thought about whether Cody wants Morgan to finish the block tower. The people agreed: They both said, "I think Cody *wants* Morgan to finish the block tower."

Check 2: Can you remind me, did they agree or disagree about what Cody wants? [child responds] Yeah, they agreed: They both said, "I think Cody *wants* Morgan to finish the block tower."

Based on this, what did Cody say to Morgan?

1. 1."Please finish my block tower." (direct statement).
2. "My block tower is almost finished." (ambiguous utterance).

Test: So, which did Cody say to Morgan: 1 or 2? [child responds].

Story 3: Snowman, A/D. Avery and Jordan were in the art room. Avery said something to Jordan about the snowman.

Other people heard what Avery said, and thought about whether Avery wants Jordan to draw on the snowman. The people agreed: They both said, "I think Avery *wants* Jordan to draw on the snowman."

Check 1: Can you remind me, did they agree or disagree about what Avery wants? [child responds] Yeah, they agreed: They both said, "I think Avery *wants* Jordan to draw on the snowman."

Based on this, what did Avery say to Jordan?

1. "Please draw on my snowman." (direct request)
2. "My snowman would look better if it was wearing something." (ambiguous utterance)

Test: So, which did Avery say to Jordan: 1 or 2? [child responds].

On a different day, Avery and Jordan were in the art room. Avery said something to Jordan about the snowman.

Other people heard what Avery said, and thought about whether Avery wants Jordan to draw on the snowman. The people disagreed: One said, "I think Avery *wants* Jordan to draw on the snowman, but the other said, 'Actually, I *don't* think Avery wants that."'

Check 2: Can you remind me, did they agree or disagree about what Avery wants? [child responds] Yeah, they disagreed: One said, "I think Avery *wants* Jordan to draw on the snowman," but the other said, "Actually I *don't* think Avery wants that." Based on this, what did Avery say to Jordan?

1. "Please draw on my snowman." (direct request)
2. "My snowman would look better if it was wearing something." (ambiguous utterance)

Test: So, which did Avery say to Jordan: 1 or 2? [child responds].

Story 4: Dog, D/A. Bailey and Devin were at the dog park. Bailey said something to Devin about the dog.

Other people heard what Bailey said, and thought about whether Bailey wants Devin to feed the dog some treats. The people disagreed: One said, "I think Bailey *wants* Devin to feed the dog some treats, but the other said, 'Actually, I *don't* think Bailey wants that."' Check 1: Can you remind me, did they agree or disagree about what Bailey wants? [child responds] Yeah, they disagreed: One said, "I think Bailey *wants* Devin to feed the dog some treats, but the other said, 'Actually, I *don't* think Bailey wants that."' Based on this, what did Bailey say to Devin?

1. "Please feed my dog some treats." (direct request)
2. "My dog eats so many treats." (ambiguous utterance)

Test: So, which did Bailey say to Devin: 1 or 2? [child responds].

On a different day, Bailey and Devin were at the dog park. Bailey said something to Devin about the dog.

Other people heard what Bailey said, and thought about whether Bailey wants Devin to feed the dog some treats. The people agreed: They both said, "I think Bailey *wants* Devin to feed the dog some treats."

Check 2: Can you remind me, did they agree or disagree about what Bailey wants? [child responds] Yeah, they agreed: They both said, "I think Bailey *wants* Devin to feed the dog some treats."

Based on this, what did Bailey say to Devin?

1. "Please feed my dog some treats." (direct request)
2. "My dog eats so many treats." (ambiguous utterance)

Test: So, which did Bailey say to Devin: 1 or 2? [child responds].

References

Ackerman, B. P. (1978). Children's understanding of speech acts in unconventional directive frames. *Child Development, 49*(2), 311–318.

Aguert, M., & Laval, V. (2013). Request complexity is no more a problem when the requests are ironic. *Pragmatics & Cognition, 21*(2), 329–339. Retrieved 2023-05-07, from https://doi.org/10.1075/pc.21.2.04agu.

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., … Iannone, R. (2023). *rmarkdown: Dynamic documents for r [Computer software manual]*. Retrieved from https://github.com/rstudio/rmarkdown (R package version 2.21).

Amemiya, J., Walker, C. M., & Heyman, G. D. (2021). Childrens developing ability to resolve disagreements by integrating perspectives. *Child Development, 92*(6). https://doi.org/10.1111/cdev.13603. Retrieved 2022-04-26, from.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bazhydai, M., Westermann, G., & Parise, E. (2020). I don't know but I know who to ask: 12montholds actively seek information from knowledgeable adults. *Developmental Science, 23*(5). https://doi.org/10.1111/desc.12938. Retrieved 2021-04-01, from.

Beck, S. R., McColgan, K. L., Robinson, E. J., & Rowley, M. G. (2011). Imagining what might be: Why children underestimate uncertainty. *Journal of Experimental Child Psychology, 110*(4), 603–610. Retrieved 2021-04-01, from https://doi.org/10.1016/j.jecp.2011.06.010.

Beck, S. R., Robinson, A. N., Ahmed, S., & Abid, R. (2011). Children's understanding that ambiguous figures have multiple interpretations. *European Journal of Developmental Psychology, 8*(4), 403–422. Retrieved 2021-04-01, from https://doi.org/10.1080/17405629.2010.515885.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function: Development of executive functions. *Child Development, 81*(6), 1641–1660. Retrieved 2023-08-10, from https://doi.org/10.1111/j.1467-8624.2010.01499.x.

Bolker, B., & Robinson, D. (2022). *broom.mixed: Tidying methods for mixed models [Computer software manual]*. Retrieved from https://CRAN.R-project.org/package=broom.mixed (R package version 0.2.9.4).

Boseovski, J. J., & Thurman, S. L. (2014). Evaluating and approaching a strange animal: Children's Trust in informant testimony. *Child Development, 85*(2), 824–834. Retrieved 2021-04-01, from https://doi.org/10.1111/cdev.12156.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*(4), 1686. Retrieved 2021-04-01, from https://doi.org/10.2307/1131725.

Carrell, P. L. (1981). Children's understanding of indirect requests: Comparing child and adult comprehension. *Journal of Child Language, 8*(2), 329–345. Retrieved 2023-04-15, from https://doi.org/10.1017/S0305000900003226.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356–373. Retrieved 2021-04-01, from https://doi.org/10.1080/15248371003700015.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *xtable: Export tables to latex or html [Computer software manual]*. Retrieved from https://CRAN.R-project.org/package=xtable (R package version 1.8–4).

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics, 9*, 519–540.

Elrod, M. M. (1987). Children's understanding of indirect requests. *The Journal of Genetic Psychology, 148*(1), 63–70. Retrieved 2023-04-15, from https://doi.org/10.1080/00221325.1987.9914537.

Foushee, R., & Srinivasan, M. (2017). Could both be right? Childrens and adults sensitivity to subjectivity in language. In *Proceedings of the 39th annual meeting of the cognitive science society*.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Frick, H., Chow, F., Kuhn, M., Mahoney, M., Silge, J., & Wickham, H. (2022). *rsample: General resampling infrastructure [Computer software manual]*. Retrieved from https://CRAN.R-project.org/package=rsample(R package version 1.1.1).

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J.. What happened? Reconstructing the past through vision and sound (preprint). PsyArXiv. Retrieved 2022-05-29, from https://osf.io/tfjdk.

Giron, A. P., Ciranka, S., Schulz, E., Van Den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2023). Developmental changes in exploration resemble stochastic optimization. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01662-1. Retrieved 2023-08-20, from.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences, 20*(11), 818–829.

Gopnik, A., & Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science, 4*(2), 175–183. Retrieved 2021-04-01, from https://doi.org/10.1111/1467-7687.00163.

Gweon, H., & Asaba, M. (2018). Order matters: Children's evaluation of underinformative teachers depends on context. *Child Development, 89*(3), e278–e292. Retrieved 2021-04-01, from https://doi.org/10.1111/cdev.12825.

Hamner, B., & Frasco, M. (2018). *Metrics: Evaluation metrics for machine learning [Computer software manual]*. Retrieved from https://CRAN.R-project.org/package=Metrics (R package version 0.1.4).

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology, 69*(1), 251–273. Retrieved 2021-04-01, from https://doi.org/10.1146/annurev-psych-122216-011710.

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development, 87*(3), 807–819. Retrieved 2021-04-01, from https://doi.org/10.1111/cdev.12527.

Houlihan, S. D., Kleiman-Weiner, M., Hewitt, L. B., Tenenbaum, J. B., & Saxe, R. (2023). Emotion prediction as computation over a generative theory of mind. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 381*(2251), 20220047. Retrieved 2023-06-25, from https://doi.org/10.1098/rsta.2022.0047.

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*(8), 589–604. Retrieved 2021-04-01, from https://doi.org/10.1016/j.tics.2016.05.011.

Kazak, A. E. (2018). *Editorial: Journal article reporting standards*. American Psychologist.

Koenig, M. A., Tiberius, V., & Hamlin, J. K. (2019). Childrens judgments of epistemic and moral agents: From situations to intentions. *Perspectives on Psychological Science, 14*(3), 344–360. Retrieved 2021-04-01, from https://doi.org/10.1177/1745691618805452.

Lane, J. D., & Harris, P. L. (2015). The roles of intuition and informants expertise in children's epistemic trust. *Child Development, 86*(3), 919–926. Retrieved 2021-04-01, from https://doi.org/10.1111/cdev.12324.

Langenhoff, A. F., Engelmann, J. M., & Srinivasan, M. (2023). Children's developing ability to adjust their beliefs reasonably in light of disagreement. *Child Development, 94*(1), 44–59. Retrieved 2023-01-04, from https://doi.org/10.1111/cdev.13838.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. John Wiley.

Ma, F., Gu, X., Tang, L., Luo, X., Compton, B. J., & Heyman, G. D. (2023). If they won't know, I won't wait: Anticipated social consequences drive children's performance on self-control tasks. *Psychological Science, 34*(11), 1220–1228. https://doi.org/10.1177/09567976231198194

Marocchini, E., Di Paola, S., Mazzaggio, G., & Domaneschi, F. (2022). Understanding indirect requests for information in high-functioning autism. *Cognitive Processing, 23*(1), 129–153. https://doi.org/10.1007/s10339-021-01056-z

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science, 24*(4). https://doi.org/10.1111/desc.13095. Retrieved 2023-06-25, from.

Mills, C. M., Al-Jabari, R. M., & Archacki, M. A. (2012). Why do people disagree? Explaining and endorsing the possibility of partiality in judgments. *Journal of Cognition and Development, 13*(1), 111–136. Retrieved 2021-04-01, from https://doi.org/10.1080/15248372.2010.547236.

Mills, C. M., & Grant, M. G. (2009). Biased decision-making: Developing an understanding of how positive and negative relationships may skew judgments. *Developmental Science, 12*(5), 784–797. Retrieved 2021-04-01, from https://doi.org/10.1111/j.1467-7687.2009.00836.x.

Mills, C. M., & Keil, F. C. (2008). Childrens developing notions of (im)partiality. *Cognition, 107*(2), 528–551. Retrieved 2021-04-01, from https://doi.org/10.1016/j.cognition.2007.11.003.

Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development, 83*(4), 1400–1415. https://doi.org/10.1111/j.1467–8624.2012.01762.x

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze. *Psychological Science, 19*(3), 211–215. https://doi.org/10.1111/j.1467–9280.2008.02069.x

Outa, J., Zhou, X. J., Gweon, H., & Gerstenberg, T. (2022). Stop, children whats that sound? Multi-modal inference through mental simulation. In *Proceedings of the 44th annual meeting of the cognitive science society*.

Plumert, J. M. (1996). Young children's ability to detect ambiguity in descriptions of location. *Cognitive Development, 11*(3), 1400–1415. https://doi.org/10.1016/S0885-2014(96) 90010-6

Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The developmental origins of selective social learning. *Current Directions in Psychological Science, 25*(1), 60–64. Retrieved 2021-04-01, from https://doi.org/10.1177/0963721415613962.

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/.

Ronfard, S., Bartz, D. T., Cheng, L., Chen, X., & Harris, P. L. (2018). Children's developing ideas about knowledge and its acquisition. In *, Vol. 54. Advances in child development and behavior* (pp. 123–151). Elsevier. https://doi.org/10.1016/bs.acdb.2017.10.005. Retrieved 2021-04-01, from.

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind, 1*(1), 4–14. Retrieved 2021-06-09, from https://doi.org/10.1162/OPMI_a_00002.

Skordos, D., & Papafragou, A. (2016). Childrens derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153*, 6–18. Retrieved 2021-04-01, from https://doi.org/10.1016/j.cognition.2016.04.006.

Sperber, D., & Wilson, D. (1987). Précis of relevance: Communication and cognition. *Behavioral and Brain Sciences, 10*(04), 697. Retrieved 2021-04-06, from https://doi.org/10.1017/S0140525X00055345.

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge University Press.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Van Der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632–638. Retrieved 2023-08-20, from https://doi.org/10.1177/1745691612463078.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Seidel, D. (2022). *scales: Scale functions for visualization [Computer software manual]*. Retrieved from https://CRAN.R-project.org/package=scales (R package version 1.2.1).

Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman and Hall/CRC (ISBN 978-1466561595).

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://yihui.org/knitr/ (ISBN 9781498716963).

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://bookdown.org/ yihui/bookdown (ISBN 978-1138700109).

Xie, Y. (2023a). *bookdown: Authoring books and technical documents with r markdown [Computer software manual]*. Retrieved from https://github.com/rstudio/ bookdown (R package version 0.33).
Xie, Y. (2023b). *knitr: A general-purpose package for dynamic report generation in r [Computer software manual]*. Retrieved from https://yihui.org/knitr/ (R package version 1.42).
Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://bookdown.org/ yihui/ rmarkdown.
Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://bookdown.org/yihui/rmarkdown -cookbook.
Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax [Computer software manual]*. Retrieved from https://CRAN.R-project.org/ package=kableExtra (R package version 1.3.4).