1   Predicting pragmatic cue integration in adults' and children's inferences about novel word

2   meanings

3   Manuel Bohn[1,2,3], Michael Henry Tessler[4], Megan Merrick[1], & Michael C. Frank[1]

4   [1] Department of Psychology, Stanford University

5   [2] Leipzig Research Center for Early Child Development, Leipzig University

6   [3] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary

7   Anthropology

8   [4] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

9   Author Note

10   Correspondence concerning this article should be addressed to Manuel Bohn, Max

11   Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany.

12   E-mail: manuel_bohn@eva.mpg.de

Abstract

Language is learned in complex social settings where listeners must reconstruct speakers'
intended meanings from context. To navigate this challenge, children can use pragmatic
reasoning to learn the meaning of unfamiliar words. One important challenge for pragmatic
reasoning is that it requires integrating multiple information sources. Here we study this
integration process. We isolate two sources of pragmatic information (common ground and
expectations about informativeness) and – using a probabilistic model of conversational
reasoning – formalize how they should be combined and how this process might develop. We
use this model to generate quantitative predictions, which we test against new behavioral
data from three- to five-year-old children (N = 243) and adults (N = 694). Results show
close numerical alignment between model predictions and data. Furthermore, the model
provided a better explanation of the data compared to simpler alternative models assuming
that children selectively ignore one information source. This work integrates distinct sets of
findings regarding early language and suggests that pragmatic reasoning models can provide
a quantitative framework for understanding developmental changes in language learning.

Keywords: language acquisition, social cognition, pragmatics, Bayesian modeling,
common ground

Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings

Introduction

What someone means by an utterance is oftentimes not reducible to the words they used. It takes pragmatic inference – context-sensitive reasoning about the speaker's intentions - to recover the intended meaning (Grice, 1991; Levinson, 2000; Sperber & Wilson, 2001). Contextual information comes in many forms. On the one hand, there is information provided by the utterance[1] itself. Competent language users expect each other to communicate in a cooperative way such that speakers produce utterances that are relevant and informative. Thus, semantic ambiguity can be resolved by reasoning about why the speaker produced this particular utterance (Clark, 1996; Grice, 1991; Sperber & Wilson, 2001; Tomasello, 2008). On the other hand, there is information provided by common ground (the body of mutually shared knowledge and beliefs between interlocutors)(Bohn & Koymen, 2018; Clark, 2015, 1996). Because utterances are embedded in common ground, pragmatic reasoning in context always requires information integration. But how does integration proceed? And how does it develop? Verbal theories assume that information is integrated and that this process develops but do not specify how. We bridge this gap by formalizing information integration and development in a probabilistic model of pragmatic reasoning.

Children learning their first language make inferences about intended meanings based on utterance-level and common-ground information both for language understanding and language learning (Bohn & Frank, 2019; Clark, 2009; Tomasello, 2008). Starting very early, infants expect adults to produce utterances in a cooperative way (Behne, Carpenter, &

---

[1]We use the terms utterance, utterance-level information or utterance-level cues to capture all cues that the speaker provides for their intended meaning. This includes direct referential information in the form of pointing or gazing, semantic information in the form of conventional word meanings as well as pragmatic inferences that are licenced by the particular choice of words or actions.

52 Tomasello, 2005), and expect language to be carrying information (Vouloumanos, Onishi, &

53 Pogue, 2012). By age two, children are sensitive to the informativeness of communication

54 (O'Neill & Topolovec, 2001). By age three children can use this expectation to make

55 pragmatic inferences (Stiller, Goodman, & Frank, 2015; Yoon & Frank, 2019) and to infer

56 novel word meanings (Frank & Goodman, 2014). And although older children continue to

57 struggle with some complex pragmatic inferences until age five and beyond (Noveck, 2001),

58 an emerging consensus identifies these difficulties as stemming from difficulties reasoning

59 about linguistic alternatives rather than pragmatic deficits (Barner, Brooks, & Bale, 2011;

60 Horowitz, Schneider, & Frank, 2018; Skordos & Papafragou, 2016). Thus, children's ability

61 to reason about utterance-level pragmatics is present at least by ages three to five, and

62 possibly substantially younger.

63   Evidence for the use of common ground[2] information by young children is even

64 stronger: Common ground information guides how infants produce non-verbal gestures and

65 interpret ambiguous utterances (Bohn et al., 2018; Saylor, Ganea, & Vázquez, 2011). For

66 slightly older children, common ground – in the form of knowledge about discourse novelty,

67 preferences, and even discourse expectations – also facilitates word learning (Akhtar,

68 Carpenter, & Tomasello, 1996; Bohn, Le, Peloquin, Koymen, & Frank, 2020; Saylor,

---

[2]Common ground has traditionally been defined in recursive terms: in order to be part of common ground, some piece of information has to be not just known to both interlocutors but also known to both to be shared between them (Clark, 1996). Numerous studies probed the role of sharedness of information and found that it plays a critical role in communicative interactions (Brown-Schmidt, 2009; Hanna, Tanenhaus, & Trueswell, 2003; Mozuraitis, Chambers, & Daneman, 2015). Based on this literature, one might argue that the term common ground should be restricted to describe situations in which the sharedness aspect is directly tested. However, most of this work is focused on online perspective taking. In this paper, we use the term common ground to refer to shared information that is built up over the course of an interaction - something that is supposedly easier for children (Matthews, Lieven, Theakston, & Tomasello, 2006). We assume that the consequence of a direct interaction (with matching perspectives) between the speaker and the listener is that information is not just known to both interlocutors but also assumed to be shared between them (Bohn & Koymen, 2018).

⁶⁹ Sabbagh, Fortuna, & Troseth, 2009; Sullivan, Boucher, Kiefer, Williams, & Barner, 2019).

⁷⁰ These examples, however, highlight children's use of a single pragmatic information
⁷¹ source or cue. Harnessing multiple – potentially competing – pragmatic cues poses a
⁷² separate challenge. One aspect of this integration problem is how to balance common ground
⁷³ information that is built up over the course of an interaction against information gleaned
⁷⁴ from the current utterance. Much less is known about whether and how children combine
⁷⁵ these types of information. Developmental studies that look at the integration of multiple
⁷⁶ information sources more generally find that children are sensitive to multiple sources from
⁷⁷ early on (Ganea & Saylor, 2007; Graham, San Juan, & Khu, 2017; Grosse, Moll, &
⁷⁸ Tomasello, 2010; Khu, Chambers, & Graham, 2020; Matthews et al., 2006; Nilsen, Graham,
⁷⁹ & Pettigrew, 2009). For example, in a classic study, Nadig and Sedivy (2002) found that
⁸⁰ children rapidly integrate information provided in an utterance (a particular referring
⁸¹ expression) with the speaker's perspective (the objects the speaker can see). However, the
⁸² information sources to be integrated in these studies are not all pragmatic in nature.
⁸³ Children's ability to pick out a referent following a noun reflects their linguistic knowledge
⁸⁴ and not necessarily their ability to reason about the speaker's intention in context. As a
⁸⁵ consequence, this work does not speak to the question of how and if listeners integrate
⁸⁶ different forms of *pragmatic* information. Thus, while many theories of pragmatic reasoning
⁸⁷ presuppose that pragmatic information sources are integrated, the nature of their
⁸⁸ relationship has typically not been specified.

⁸⁹ Recent innovations in probabilistic models of pragmatic reasoning provide a
⁹⁰ quantitative method for addressing the problem of integrating multiple sources of contextual
⁹¹ information. This class of computational models, which are referred to as Rational Speech
⁹² Act (RSA) models (Frank & Goodman, 2012; Goodman & Frank, 2016) formalize the
⁹³ problem of language understanding as a special case of Bayesian social reasoning. A listener
⁹⁴ interprets an utterance by assuming it was produced by a cooperative speaker who had the

goal to be informative. Being informative is defined as providing a message that would
increase the probability of the listener recovering the speaker's intended meaning in context.
This notion of contextual informativeness captures the Gricean idea of cooperation between
speaker and listener, and provides a first approximation to what we have described above as
utterance-level pragmatic information.

RSA models capture common ground information as a shared prior distribution over
possible intended meanings. Thus, a natural locus for information integration within
probabilistic models of pragmatic reasoning is the trade off between the prior probability of a
meaning and the informativeness of the utterance. This trade off between contextual factors
during word learning is a unique aspect that is not addressed by other computational models
of word learning, which have focused on learning from cross-situational, co-occurrence
statistics (Fazly, Alishahi, & Stevenson, 2010; Frank, Goodman, & Tenenbaum, 2009) or
describing generalizations about word meaning (Xu & Tenenbaum, 2007).

We make use of this framework to study pragmatic cue integration across development.
To this end, we adapt a method used in perceptual cue integration studies (Ernst & Banks,
2002): we make independent measurements of each cue's strength and then combine them
using the RSA model described above to make independent predictions about conditions in
which they either coincide or conflict. Finally, we pre-register these quantitative predictions
and test them against new data from adults and children.

We start by replicating previous findings with adults showing that listeners make
pragmatic inferences based on non-linguistic properties of utterances in isolation (experiment
1). Then we show that adults make inferences based on common ground information
(experiment 2A and 2B). We use data from these experiments as parameters to generate a
priori predictions from RSA models about how utterance and common ground information
should be integrated. We consider three models that make different assumptions about the
integration process: In the *integration model*, the two information sources are integrated with

121 one another. The other two models are lesion models that assume that participants focus on

122 one type of information and disregard the other whenever they are presented together.

123 According to the *no common ground* model, participants focus only on the utterance

124 information and in the *no informativeness* model, only common ground information is

125 considered. We compare predictions from these models to new empirical data from

126 experiments in which utterance and common ground information are manipulated

127 simultaneously (Experiment 3 and 4).

128 After successfully validating this approach with adults in study 1, we apply the same

129 model-driven experimental procedure to children (study 2): We first show that they make

130 pragmatic inferences based on utterance and common ground information separately

131 (experiment 5 and 6). Then we generate a priori model predictions and compare them to data

132 from an experiment in which both information sources have to be integrated (experiment 7).

133 Taken together, this work makes two primary contributions: first, it shows that both

134 adults and children integrate utterance-level and common-ground information flexibly.

135 Second, it uses Bayesian data analysis within the RSA framework to provide a model for

136 understanding the multiple loci for developmental change in complex behaviors like

137 contextual communication.

138 **Study 1: Adults**

139 **Participants**

140 Adult participants were recruited via Amazon Mechanical Turk (MTurk) and received

141 payment equivalent to an hourly wage of ~ \$9. Each participant contributed data to only

142 one experiment. Experiment 1 and each manipulation of experiment 2 had $N = 40$

143 participants. Sample size in experiment 3 was $N = 121$. $N = 167$ participated in the

144 experiments to measure the strong, medium and weak preference and novelty manipulations
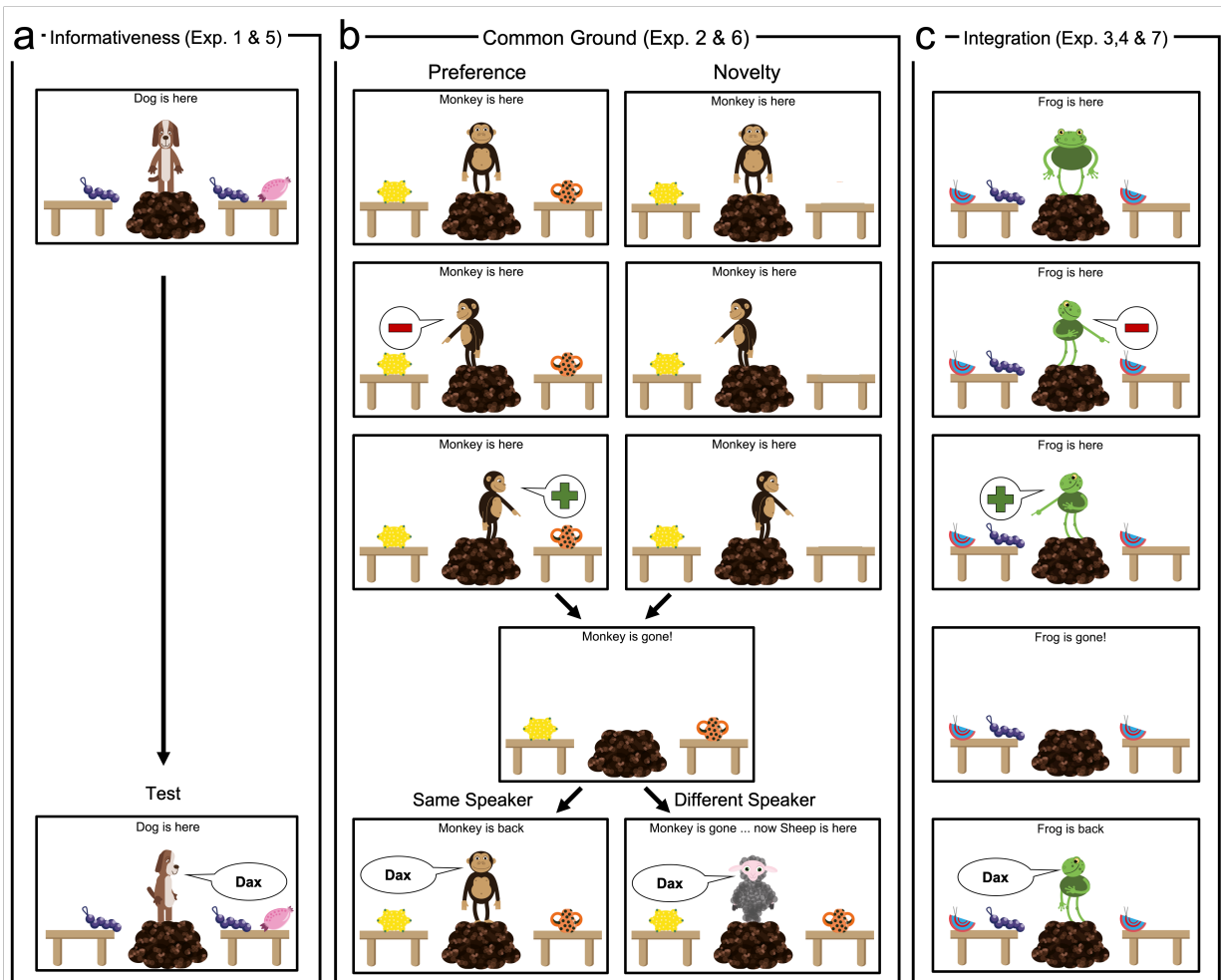
Figure 1. Schematic experimental procedure with screenshots from the adult experiments. In all conditions, at test (bottom), the speaker ambiguously requested an object using a non-word (e.g. "dax"). Participants clicked on the object they thought the speaker referred to. Speech bubbles represent pre-recorded utterances. Informativeness (a) translated to making one object less frequent in context. Common ground (b) was manipulated by making one object preferred by or new to the speaker. Green plus signs represent utterances that expressed preference and red minus signs represent utterances that expressed dispreference (see main text for details). Integration (c) combined informativeness and common ground manipulations. One integration condition is shown here: preference - same speaker - congruent.

₁₄₅ that went into experiment 4. Finally, experiment 4 had $N = 286$ participants. Sample sizes

₁₄₆ in all adult experiments were chosen to yield at least 120 data points per cell. All studies

₁₄₇ were approved by the Stanford Institutional Review Board (protocol no. 19960).

Materials

₁₄₉ All experimental procedures were pre-registered (see

₁₅₀ https://osf.io/u7kxe/registrations). Experimental stimuli are freely available in the following

₁₅₁ online repository: https://github.com/manuelbohn/mcc. All experiments were framed as

₁₅₂ games in which participants would learn words from animals. They were implemented in

₁₅₃ HTML/JavaScript as a website. Adults were directed to the website via MTurk and

₁₅₄ responded by clicking objects. For each animal character, we recorded a set of utterances

₁₅₅ (one native English speaker per animal) that were used to provide information and make

₁₅₆ requests. All experiments started with an introduction to the animals and two training trials

₁₅₇ in which familiar objects were requested (car and ball). Subsequent test trials in each

₁₅₈ condition were presented in a random order.

Analytic approach

₁₆₀ We preregistered sample sizes, inferential statistical analysis and computational models

₁₆₁ for all experiments. All deviations from the registered analysis plan are explicitly mentioned.

₁₆₂ All analyses were run in R (R Core Team, 2018). All p-values are based on two sided

₁₆₃ analysis. Cohen's d (computed via the function `cohensD`) was used as effect size for t-tests.

₁₆₄ Frequentist logistic GLMMs were fit via the function `glmer` from the package `lme4` (Bates,

₁₆₅ Mächler, Bolker, & Walker, 2015) and had a maximal random effect structure conditional on

₁₆₆ model convergence. Details about GLMMs including model formulas for each experiment

₁₆₇ can be found in the Supplementary Material available online. Probabilistic models and

model comparisons were implemented in WebPPL (Goodman & Stuhlmüller, 2014) using the

R package `rwebppl` (Braginsky, Tessler, & Hawkins, 2019). In experiment 3, 4 and 7, we

compared probabilistic models based on Bayes Factors which were calculated from the

marginal likelihoods of each model given the data. Details on models, including information

about priors for parameter estimation and Markov chain Monte Carlo settings can be found

in the Supplementary Material available online. Code to run the models is available in the

associated online repository.

Experiment 1

Methods.   In experiment 1, participants could learn which object a novel word

referred to by assuming that the speaker communicated in an informative way (Frank &

Goodman, 2014). The speaker was located between two tables, one with two novel objects, A

and B, and the other with only object A (Fig 1a). At test, the speaker turned and pointed

to the table with the two objects (A and B) and used a novel word to request one of them.

The same utterance was used to make a request in all adult studies ("Oh cool, there is a

[non-word] on the table, how neat, can you give me the [non-word]?"). Participants could

infer that the word referred to object B via the counter-factual inferences that, if the

(informative) speaker had wanted to refer to object A, they would have pointed to the table

with the single object (this being the least ambiguous way to refer to that object). In the

control condition, both tables contained both objects and no inference could be made based

on the speaker's behavior. Participants received six trials, three per condition.

Results.   Participants selected object B above chance in the test condition (mean =

0.74, 95% CI of mean = [0.65; 0.83], t(39) = 5.51, $p < .001$, d = 0.87) and more often

compared to the control condition ($\beta = 1.28$, se = 0.29, $p < .001$, see Fig 2). This finding

replicates earlier work showing that adult listeners expect speakers to communicate in an

informative way.

Experiment 2

193

194   **Methods.**   In experiments 2A and 2B, we tested if participants use common ground

195   information that is specific to a speaker to identify the referent of a novel word (Akhtar et

196   al., 1996; Diesendruck, Markson, Akhtar, & Reudor, 2004; Saylor et al., 2009). In experiment

197   2A, the speaker expressed a preference for one of two objects (Fig 1b, left). The animal

198   introduced themselves, then turned to one of the tables and expressed either that they liked

199   ("Oh wow, I really like that one") or disliked ("Oh bleh, I really don't like that one") the

200   object before turning to the other side and expressing the respective other attitude. Next the

201   animal disappeared and, after a short pause, either the same or a different animal returned

202   and requested an object while facing straight ahead. Participants could use the speakers

203   preference to identify the referent when the same speaker returned but not when a different

204   speaker appeared whose preferences were unknown.

205   In experiment 2B, common ground information came in the form of novelty (Fig 1b,

206   right). The animal turned to one of the sides and commented either on the presence ("Aha,

207   look at that") or the absence ("Hm. . . , nothing there") of an object before turning to the

208   other side and commenting in a complementary way. Later, a second object appeared on the

209   previously empty table. Then the speaker used a novel word to request one of the objects.

210   The referent of the novel word could be identified by assuming that the speaker uses it to

211   refer to the object that is new to them. This inference was not licensed when a different

212   speaker returned to whom both objects were equally new. For both novelty and preference,

213   participants received six trials, three with the same and three with the different speaker.

214   **Results.**   In experiment 2A, participants selected the preferred object above chance

215   (mean = 0.97, 95% CI of mean = [0.93; 1], t(39) = 29.14, $p < .001$, d = 4.61) and more so

216   than in the speaker change control condition ($\beta = 2.92$, se = 0.57, $p < .001$).

217   In experiment 2B, participants selected the novel object above chance (mean = 0.83,

218 95% CI of mean = [0.73; 0.93], t(39) = 6.77, $p < .001$, d = 1.07) when the same speaker

219 made the request and more often compared to when a different speaker made the request (β

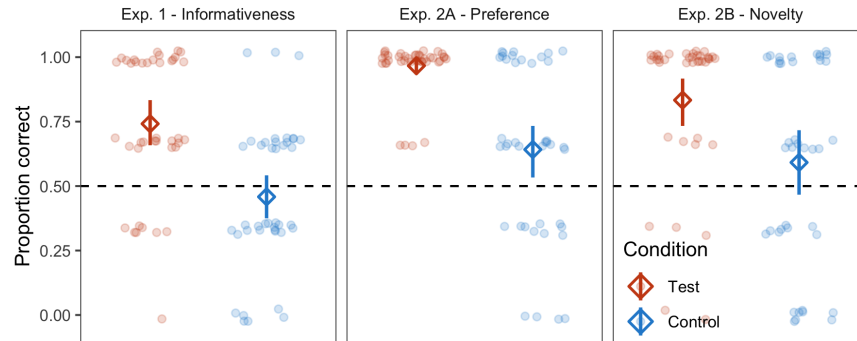220 = 6.27, se = 1.96, $p = .001$, see Fig 2).

Figure 2. Results from experiments 1, 2A, and 2B for adults. For preference and novelty, control refers to a different speaker (see Fig 1b). Transparent dots show data from individual participants (slightly jittered to avoid overplotting), diamonds represent condition means, error bars are 95% CIs. Dashed line indicates performance expected by chance.

221 **Modelling information integration**

222   Experiments 1 and 2 confirmed that adults make pragmatic inferences based on

223 information provided by the utterance as well as by common ground and provided

224 quantitative estimates of the strength of these inferences for use in our model. We modeled

225 the integration of utterance informativity and common ground as a process of socially-guided

226 probabilistic inference, using the results of experiments 1 and 2 to inform key parameters of

227 a computational model. The Rational Speech Act (RSA) model architecture introduced by

228 Frank and Goodman (2012) encodes conversational reasoning through the perspective of a

229 listener ("he" pronoun) who is trying to decide on the intended meaning of the utterance he

230 heard from the speaker ("she" pronoun). The basic idea is that the listener combines his

231 uncertainty about the speaker's intended meaning - a prior distribution over referents P(r) -

232   with his generative model of how the utterance was produced: a speaker trying to convey

233   information to him. To adapt this model to the word learning context, we enrich this basic

234   architecture with a mechanism for expressing uncertainty about the meanings of words

235   (lexical uncertainty) - a prior distribution over lexica P(L) (Bergen, Levy, & Goodman, 2016).

$$P_L(r, \mathcal{L}|u) \propto P_S(u|r, \mathcal{L}) \cdot P(\mathcal{L}) \cdot P(r)$$

236   In the above equation, the listener is trying to jointly resolve the speaker's intended

237   referent r and the meaning of words (thus learning the lexicon $\mathcal{L}$). He does this by imagining

238   what a rational speaker would say, given the referent they are trying to communicate and a

239   lexicon. The speaker is an approximately rational Bayesian actor (with degree of rationality

240   alpha), who produces utterances as a function of their informativity. The space of utterances

241   the speaker could produce depends upon the lexicon $P(u|\mathcal{L})$; simply put, the speaker labels

242   objects with the true labels under a given lexicon L (see Supplementary Material available

243   online for details):

$$P_S(u|r, \mathcal{L}) \propto Informativity(u; r)^{\alpha} \cdot P(u|\mathcal{L})$$

244   The informativity of an utterance for a referent is taken to be the probability with

245   which a naive listener, who only interprets utterances according to their literal semantics,

246   would select a particular referent given an utterance.

$$Informativity(u; r) = P(r|u) \propto P(r) \cdot \mathcal{L}_{point}$$

247   The speaker's possible utterances are pairs of linguistic and non-linguistic signals,

248   namely labels and points. Because the listener does not know the lexicon, the informativity

249   of an utterance comes from the speaker's point, the meaning of which is encoded in $\mathcal{L}_{point}$

250 and is simply a truth-function checking whether or not the referent is at the location picked

251 out by the speaker's point. Though the speaker makes their communicative decision

252 assuming the listener does not know the meaning of the labels, we assume that in addition to

253 a point, the speaker produces a label consistent with their own lexicon $\mathcal{L}$, described by

254 $P(u|\mathcal{L})$ (see Supplementary Material available online for modeling details).

255   This computational model provides a natural avenue to formalize quantitatively how

256 informativeness and common ground trade-off during word learning. As mentioned above,

257 the common ground shared between speaker and listener plays the role of the listener's prior

258 distribution over meanings, or types of referents, that the speaker might be referring to and

259 which we posit depends on prior interactions around the referents in the present context

260 (e.g., preference or novelty; experiment 2A and B). We use the results from experiment 2 to

261 specify this distribution. The in-the-moment, contextual informativeness of the utterance is

262 captured in the likelihood term, whose value depends on the rationality parameter α.

263 Assumptions about rationality may change depending on context and we therefore used the

264 data from experiment 1 to specify α (see Supplementary Material available online for details

265 about these parameters).

266   The model generates predictions for situations in which utterance and common ground

267 expectations are jointly manipulated (Fig 1c - see Supplementary Material available online

268 for additional details and a worked example of how predictions were generated). In addition

269 to the parameters fit to the data from previous experiments, we include an additional noise

270 parameter, which can be thought of as reflecting the cost that comes with handling and

271 integrating multiple information sources. Technically it estimates the proportion of responses

272 better explained by a process of random guessing than by pragmatics; we estimate this

273 parameter from the observed data (experiment 3). Including the noise parameter greatly

274 improved the model fit to the data (see Supplementary Material available online for details).

275 We did not pre-register the inclusion of a noise parameter for experiment 3 but did so for all

276 subsequent experiments.

Experiment 3

278 **Methods.** In experiment 3, we combined the procedures of experiment 1 and 2A or
279 2B. The test setup was identical to experiment 1, however, before making a request, the
280 speaker interacted with the objects so that some of them were preferred by or new to them
281 (Fig 1c). This combination resulted in two ways in which the two information sources could
282 be aligned with one another. In the congruent condition, the object that was the more
283 informative referent was also the one that was preferred by or new to the speaker. In the
284 incongruent condition, the other object was the one that was preferred by or new to the
285 speaker. Taken together, there were 2 (novelty or preference) x 2 (same or different speaker)
286 x 2 (congruent or incongruent) = 8 conditions in experiment 3. For each of these eight
287 conditions, we generated model predictions using the modelling framework introduced above.
288 The test hypothesis about how information is integrated we compared the three models
289 introduced in the introduction: The *integration model* in which both information sources are
290 flexibly combined, the *no common ground model* that focused only on utterance-level
291 information and the the *no informativeness model* that focused only on common ground
292 information.

293 Participants completed eight trials for one of the common ground manipulations with
294 two trials per condition (same/different speaker x congruent/incongruent). Conditions were
295 presented in a random order. We discuss and visualize the results as the proportion with
296 which participants chose the more informative object (i.e., the object that would be the more
297 informative referent when only utterance information is considered).

298 **Results.** As a first step, we used a GLMM to test whether participants were
299 sensitive to the different ways in which information could be aligned. We found that

₃₀₀ participants distinguished between congruent and incongruent trials when the speaker

₃₀₁ remained the same (model term: `alignment x speaker`; β = -2.64, se = 0.48, $p < .001$).

₃₀₂ Thus, participants were sensitive to the different combinations of manipulations.

₃₀₃   As a second step, we compared the model predictions to the data. Participants'

₃₀₄ average responses were highly correlated with the predictions from the *integration model* in

₃₀₅ each condition (Fig 3b). When comparing model, we found that model fit was considerably

₃₀₆ better for the *integration model* compared to the *no common ground model* (Bayes Factor

₃₀₇ (BF) = 4.2e+53) or the *no informativeness model* (BF = 2.5e+34), suggesting that

₃₀₈ participants considered and integrated both sources of information.

₃₀₉   Finally, we examined the noise parameter for each model. The estimated proportion of

₃₁₀ random responses according to the *integration model* was 0.30 (95% Highest Density Interval

₃₁₁ (HDI): 0.23 - 0.36). This parameter was substantially lower for the *integration model*

₃₁₂ compared to the alternative models (*no common ground model*: 0.60 [0.46 - 0.72]; *no*

₃₁₃ *informativeness model*: 0.41 [0.33 - 0.51]), lending additional support to the conclusion that

₃₁₄ the *integration model* better captured the behavioral data. Rather than explaining

₃₁₅ systematic structure in the data, the alternative models achieved their best fit only by

₃₁₆ assuming a very high level of noise.

Experiment 4

₃₁₈   **Methods.**   To test if the *integration model* makes accurate predictions for different

₃₁₉ combinations, we first replicated and then extended the results of experiment 3 to a broader

₃₂₀ range of experimental conditions. Specifically, we manipulated the strength of the common

₃₂₁ ground information (3 levels - strong, medium and weak - for preference and 2 levels - strong

₃₂₂ and medium - for novelty) by changing the way the speaker interacted with the objects prior

₃₂₃ to the request. The procedural details and statistical analysis for these these manipulations
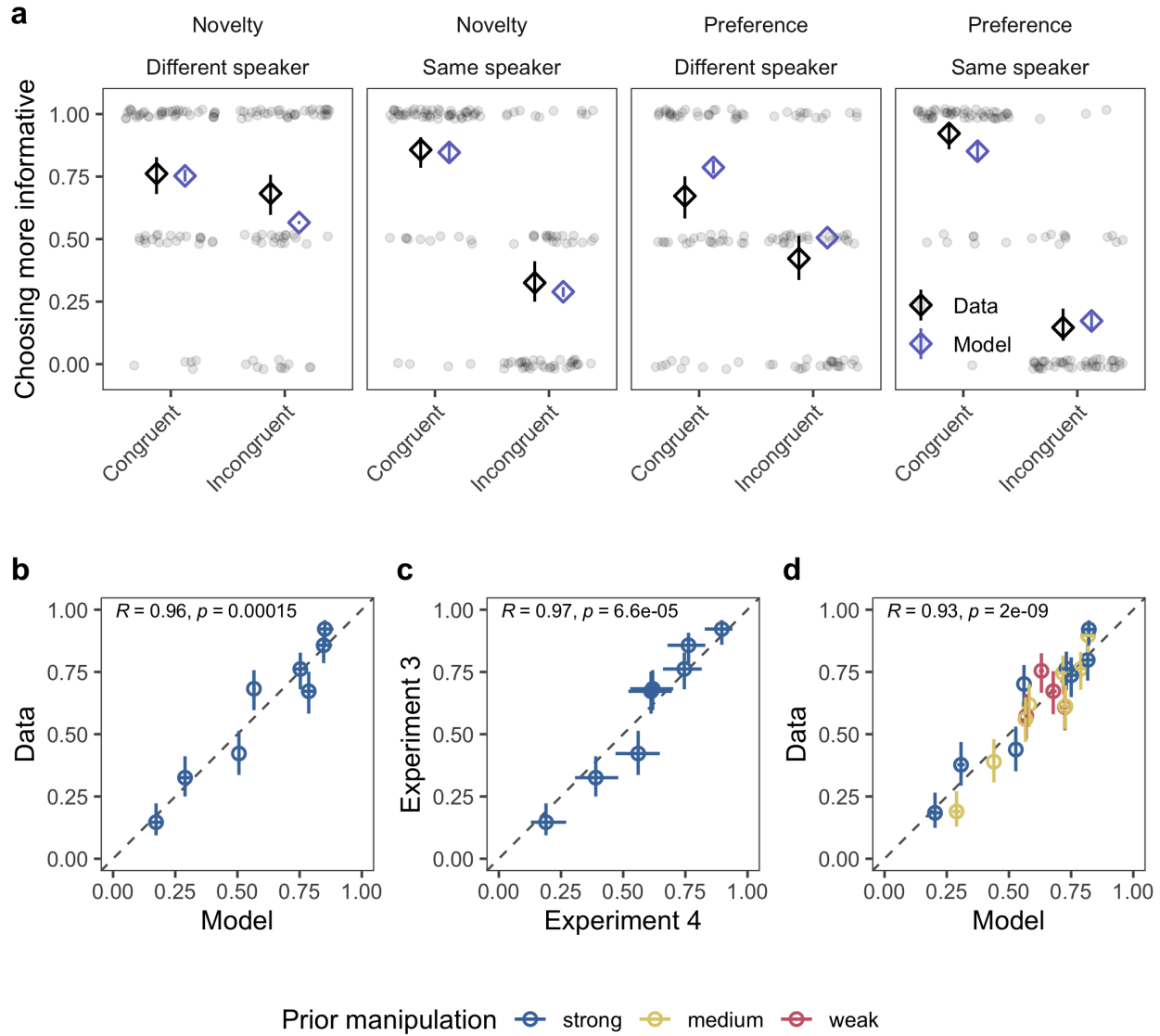
Figure 3. Results from experiment 3 and 4 for adults. Data and model predictions by condition for experiment 3 (a). Transparent dots show data from individual participants (slightly jittered to avoid overplotting), diamonds represent condition means. Correlation between model predictions and data in Experiment 3 (b), between data in Experiment 3 and the direct replication in experiment 4 (c) and between model predictions and data in experiment 4 (d). Coefficients and p-values are based on Pearson correlation statistics. Error bars represent 95% HDIs.

324 are described in the Supplementary Material available online. For experiment 4, we paired

325 each level of prior strength manipulation with the informativeness inference in the same way

326 as in experiment 3. This resulted in a total of 20 conditions, for which we generated a priori

327 model predictions in the same way as in experiment 3. That is, we conducted a separate

328 experiment for each level of prior strength and common ground manipulation to estimate the

329 prior probability of each object following this particular manipulation (analogous to

330 experiment 2). This prior distribution was then passed through the model for the congruent

331 and incongruent conditions, resulting in a unique prediction for each of the 20 condition.

332 Given the graded nature of the prior manipulations, experiment 4 basically tests how well

333 the model performs with different types of prior distributions.

334   The strong prior manipulation in experiment 4 was a direct replication of experiment 3

335 (see Fig 3c). Each participant was randomly assigned to a common ground manipulation and

336 a level of prior strength and completed eight trials in total, two in each unique condition in

337 that combination.

338   **Results.**   The direct replication of experiment 3 within experiment 4 showed a very

339 close correspondence between the two rounds of data collection (see Fig 3c). GLMM results

340 for experiment 4 can be found in the Supplementary Material available online. Here we focus

341 on the analysis based on the probabilistic models. Model predictions from the *integration*

342 *model* were again highly correlated with the average response in each condition (see Fig 3d).

343 We evaluated model fit for the same models as in experiment 3 and found again that the

344 *integration model* fit the data much better compared to the *no common ground* (BF =

345 4.7e+71) or the *no informativeness model* (BF = 8.9e+82). The inferred level of noise based

346 on the data for the *integration model* was 0.36 (95% HDI: 0.31 - 0.41), which was similar to

347 experiment 3 and again lower compared to the alternative models (*no common ground model*:

348 0.53 [0.46 - 0.62]; *no informativeness model*: 0.67 [0.59 - 0.74]).

₃₄₉ **Study 2: Children**

₃₅₀   The previous section showed that competent language users flexibly integrate

₃₅₁ information during pragmatic word learning. Do children make use of multiple information

₃₅₂ sources during word learning as well? When does this integration emerge developmentally?

₃₅₃ While many verbal theories of language learning imply that this integration takes place, the

₃₅₄ actual process has neither been described nor tested in detail. Here we provide an

₃₅₅ explanation in the form of our *integration model* and test if it is able to capture children's

₃₅₆ word learning. Embedded in the assumptions of the model is the idea that developmental

₃₅₇ change is change in the strength of the individual inferences, leading to a change in the

₃₅₈ strength of the integrated inference. As a starting point, our model assumes developmental

₃₅₉ continuity in the integration process itself (Bohn & Frank, 2019), though this assumption

₃₆₀ could be called into question by a poor model fit. The study for children followed the same

₃₆₁ general pattern as the one for adults. We generated model predictions for how information

₃₆₂ should be integrated by first measuring children's ability to use utterance (informativeness)

₃₆₃ and common ground (preference) information in isolation when making pragmatic inferences.

₃₆₄ We then adapted our model to study developmental change: We sampled children

₃₆₅ continuously between 3.0 and 5.0 years of age – a time in which children have been found to

₃₆₆ make the kind of pragmatic inferences we studied here (Bohn & Frank, 2019; Frank &

₃₆₇ Goodman, 2014) - and generated model predictions for the average developmental trajectory

₃₆₈ in each condition.

₃₆₉ **Participants**

₃₇₀   Children were recruited from the floor of the Children's Discovery Museum in San Jose,

₃₇₁ California, USA. Parents gave informed consent and provided demographic information.

₃₇₂ Each child contributed data to only one experiment. We collected data from a total of 243

₃₇₃ children between 3.0 and 5.0 years of age. We excluded 15 children due to less than 75% of

reported exposure to English, five because they responded incorrectly on 2/2 training trials, three because of equipment malfunction, and two because they quit before half of the test trials were completed. The final sample size in each experiment was as follows: $N = 62$ (41 girls, mean age = 4) in experiment 5, $N = 61$ (28 girls, mean age = 3.99) in experiment 6 and $N = 96$ (54 girls, mean age = 3.96) in experiment 7. For experiment 5 and 6, we also tested two-year-olds but did not find sufficient evidence that they use utterance and/or common ground information in the tasks we used to justify investigating their ability to integrate the two. Sample sizes in all experiments were chosen to yield at least 80 data points in each cell for each age group.

Materials

Experiments were implemented in the same general way as for adults. Children were guided through the games by an experimenter and responded by touching objects on the screen of an iPad tablet (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016).

Experiment 5

Methods.   Experiment 5 for children was modeled after Frank & Goodman (2014). Instead of on tables, objects were presented as hanging in trees (to facilitate showing points to distinct locations). After introducing themselves, the animal turned to the tree with two objects and said: "This is a tree with a [non-word], how neat, a tree with a [non-word]"). Next, the trees and the objects in them disappeared and new trees replaced them. The two objects from the tree the animal turned to previously were now spread across the two trees (one object per tree, position counterbalanced). While facing straight, the animal first said "Here are some more trees" and then asked the child to pick the tree with the object that corresponded to the novel word ("Which of these trees has a [non-word]?"). Children received six trials in a single test condition.

398   **Results.**   To compare children's performance to chance level, we binned age by year.

399   Four-year-olds selected the more informative object (i.e. the object that was unique to the

400   location the speaker turned to) above chance (mean = 0.62, 95% CI of mean = [0.53; 0.71],

401   t(29) = 2.80, p = .009, d = 0.51). Three-year-olds, on the other hand, did not (mean = 0.46,

402   95% CI of mean = [0.41; 0.52], t(31) = -1.31, p = .198, d = 0.23). Consequently, when we fit

403   a GLMM to the data with age as a continuous predictor, performance increased with age (β

404   = 0.38, se = 0.11, p < .001, see Fig 4). Thus, children's ability to use utterance information

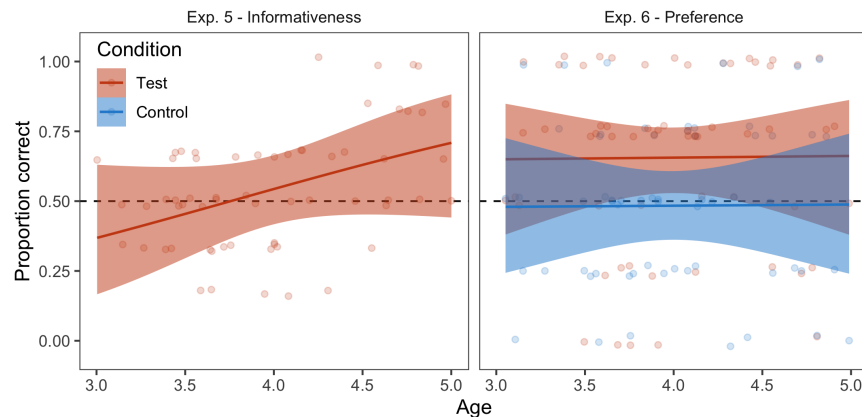405   in a word learning context increased with age.

Figure 4. Results from experiment 5 and 6 for children. For preference, control refers to to
the different speaker condition (see Fig. 1B). Transparent dots show data from individual
participants (slightly jittered to avoid overplotting), regression lines show fitted linear models
with 95% CIs. Dashed line indicates performance expected by chance.

406   **Experiment 6**

407   **Methods.**   In experiment 6, we assessed whether children use common ground

408   information to identify the referent of a novel word. We tested children only with the

409   preference manipulation[3]. The procedure for children was identical to the preference

---

[3]We initially tested children with the novelty as well as the preference manipulation. We found that
children made the basic inference in that they selected the object that was preferred by or new to the speaker,

410 manipulation for adults. Children received eight trials, four with the same and four with a

411 different speaker.

412   **Results.** Four-year-olds selected the preferred object above chance when the same

413 speaker made the request (mean = 0.71, 95% CI of mean = [0.61; 0.81], t(30) = 4.14, $p <$

414 .001, d = 0.74), whereas three-year-olds did not (mean = 0.60, 95% CI of mean = [0.47;

415 0.73], t(29) = 1.62, $p = $ .117, d = 0.30). However, when we fit a GLMM to the data with age

416 as a continuous predictor, we found an effect of speaker identity ($\beta = 0.89$, se = 0.24, $p <$

417 .001) but no effect of age ($\beta = 0.02$, se = 0.16, $p = $ .92) or interaction between speaker

418 identity and age ($\beta = $ -0.01, se = 0.23, $p = $ .97, see Fig 4). Thus, children across the age

419 range used common ground information to infer the referent of a novel word.

Modelling information integration in children

421   Model predictions for children were generated using the same model described above

422 for adults. However, to incorporate developmental change in the model, we allowed the

423 rationality parameter α and the prior distribution over objects to change with age. That is,

424 instead of a single value, we inferred the intercept and slope for each parameter that best

425 described the developmental trajectory in the data of experiment 5 and 6. These parameter

426 settings were then used to generate age sensitive model predictions in 2 (same or different

427 speaker) x 2 (congruent or incongruent) = 4 conditions. As for adults, all models included a

428 noise parameter, which was estimated based on the data of experiment 7.

Experiment 7

---

but found little evidence that children distinguished between requests made by the same speaker or a different

speaker in the case of novelty. This finding contrasts with earlier work (Diesendruck et al., 2004). since our

focus was on how children integrate informativeness and common ground, we did not follow up on this finding

but dropped the novelty manipulation and focused on preference for the remainder of the study.

⁴³⁰   **Methods.**   In experiment 7, we combined the procedures of experiment 5 and 6 and

⁴³¹ collected new data from children between 3.0 and 5.0 years of age in each of the four

⁴³² conditions (Fig 1c). We again inserted the preference manipulation into the setup of

⁴³³ experiment 5. After greeting the child, the animal turned to one of the trees, pointed to an

⁴³⁴ object (object was temporarily enlarged and moved closer to the animal) and expressed

⁴³⁵ liking or disliking. Then the animal turned to the other tree and expressed the other

⁴³⁶ attitude for the other kind of object. Next, the animal disappeared and either the same or a

⁴³⁷ different animal returned. The rest of the trial was identical to the request phase of

⁴³⁸ experiment 5. Children received eight trials, two per condition (same/different speaker x

⁴³⁹ congruent/incongruent) in a randomized order.

⁴⁴⁰   **Results.**   As a first step, we used a GLMM to test whether children were sensitive to

⁴⁴¹ the different ways in which information could be aligned. Children's propensity to

⁴⁴² differentiate between congruent and incongruent trials for the same or a different speaker

⁴⁴³ increased with age (model term: `age x alignment x speaker`; β = -0.89, se = 0.36, p =

⁴⁴⁴ .013).

⁴⁴⁵   Analyses comparing the model predictions from the probabilistic models to the data

⁴⁴⁶ suggest that children flexibly integrate both common ground and informativity information.

⁴⁴⁷ Furthermore, this integration process is accurately captured by the *integration model* at least

⁴⁴⁸ for four-year-olds. For the correlational analysis, we binned model predictions and data by

⁴⁴⁹ year. There was a substantial correlation between the predicted and measured average

⁴⁵⁰ response for four-year-olds, but less so for three-year-olds (Fig 5b). One of the reasons for

⁴⁵¹ the latter was the low variation between conditions. For the model comparison, we treated

⁴⁵² age continuously. As with adults, we found a much better model fit for the *integration model*

⁴⁵³ compared to the *no common ground* (BF = 577) or the *no informativeness model* (BF =

⁴⁵⁴ 10560).

⁴⁵⁵   The inferred level of noise based on the data for the integration model was 0.51 (95%

HDI: 0.26 - 0.77), which was lower compared to the alternative models considered (*no*

common ground model: 0.81 [0.44 - 1.00]; *no informativeness model*: 0.99 [0.88 - 1.00]) but

numerically higher than that of adults.

The high level of inferred noise moved the model predictions for children in all

conditions close to chance level. We therefore compared two additional sets of models with

different parameterizations of the noise parameter that emphasized differences between

conditions in the model predictions more (see Supplementary Material available online and

Fig 5a). This analysis was not pre-registered. Parameter free models did not include a noise

parameter and developmental noise models allowed the noise parameter to change with age.

In each case, the *integration model* provided a better fit compared to the alternative models

(*no common ground*: parameter free BF $= 334$, developmental noise BF $= 16361$; *no*

informativeness: parameter free BF $= 20$, developmental noise BF $= 1e+06$). The

developmental noise parameter for the integration model decreased with age, suggesting that

older children behaved more in line with model predictions compared to younger children

(see Fig. S13 in Supplementary Material available online).

Discussion

Integrating multiple sources of information is an integral part of human communication

(Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). To infer the intended meaning of

an utterance, listeners must combine their knowledge of communicative conventions

(semantics and syntax) with social expectations about their interlocutor. This integration is

especially vital in early language learning, and the different varieties of pragmatic

information are among the most important sources (Bohn & Frank, 2019). But how are

pragmatic cues integrated during word learning? Here we used a Bayesian cognitive model to

formalize this integration process. We studied how utterance-level (Gricean) expectations

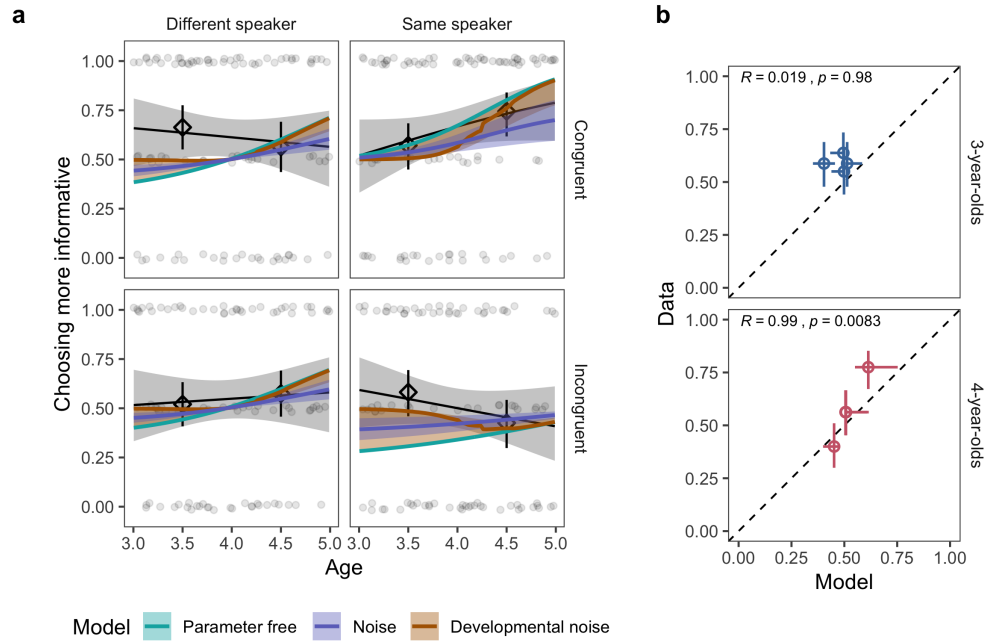about informative communication are integrated with common ground information. Adults'

Figure 5. Results from experiment 7 for children. Model predictions and data across age in the four conditions (a). Transparent black dots show data from individual participants and black lines show conditional means of the data with 95% CI. Black diamonds show the mean of the data for age bins by year and error bars show 95% CIs. Correlation between model predictions (with noise parameter) and condition means binned by year (b). Coefficients and p-values are based on Pearson correlation statistic. Error bars and shaded regions represent 95% HDIs. For 4-year-olds, two conditions yielded the same data means and model predicitons and are thus plotted on top of each other.

and children's learning was best predicted by a model in which both sources of information traded-off flexibly. Alternative models that considered only one source of information made substantially worse predictions.

All of the models we compared here integrated some explicit structure, rather than (for example) simply weighing information sources by some ratio. We made this decision because we wanted to make predictions within a framework in which the models were models of the task, rather than simply models of the data. That is, inferences are not computed separately

by the modeler and specified as inputs to a regression model, but instead are the results of an integrated process that operates over a (schematic) representation of the experimental stimuli. Further, our models are variants derived from the broader RSA framework, which has been integrated into larger systems for language learning in context (Cohn-Gordon, Goodman, & Potts, 2018; Monroe, Hawkins, Goodman, & Potts, 2017; Wang, Liang, & Manning, 2016).

How is information integrated in this context? The *integration model* assumes that the informativeness of an utterance depends on the common ground shared between interlocutors. This conception of information integration explains the seemingly counterintuitive predictions of the model. For example, one might expect that the model predicts a performance at chance level in the same speaker – incongruent conditions because the two cues "pull" the listener in opposite directions. Instead, the model predicts a performance below chance, favoring the object implicated by the prior (which also matches participants' responses). This is because the listener assumes that the speaker takes the common ground shared between the speaker and the (naive) listener as a starting point when computing the effect of each utterance. As a consequence, when prior interactions strongly implicate one object as the more likely referent, the speaker reasons that this object will be the inferred referent of any semantically plausible utterance, even when the same utterance would point to a different object in the absence of common ground. Taken together, our model advances classic theories on pragmatic language comprehension (Grice, 1991; Sperber & Wilson, 2001) and learning (Bruner, 1983; Tomasello, 2009) by providing an explicit and formal description of the integration process, thereby offering an answer to the question of *how* information may be integrated during pragmatic word learning. Predictions generated based on this process accurately captured adults' inferences across a wide range of conditions.

The *integration model* predicted information integration in four-year-olds. However, the model did not successfully describe three-year-olds' inferences; thus, it is possible that

514 they were not able to integrate information sources. But our findings are also consistent with

515 a simpler explanation, namely that the overall weaker responses we observed in the

516 independent measurement experiments (experiments 5 and 6), combined with some noise in

517 responding, led the younger children to appear relatively random in their responses. As a

518 consequence, there was not much variation in three-year-old's responses for the model to

519 explain.

520   We did not model the social-cognitive processes that specify the probability of an

521 object being the referent given common ground - we simply measured it empirically. As a

522 consequence, our approach treats common ground as equivalent to more basic manipulations

523 of contextual salience (e.g. in Frank & Goodman, 2012). Thus, our model would not

524 differentiate between a situation in which an object would be salient because it has been the

525 focus of an interaction and one in which it would be more salient because it was big or

526 colorful. A starting point for explicitly modeling common ground would be the work by

527 Heller and colleagues (Heller, Parisien, & Stevenson, 2016; Mozuraitis, Stevenson, & Heller,

528 2018). In their work, they focus on how listeners identify the referent of ambiguous referring

529 expressions. Their probabilistic model simultaneously considers the (differing) perspectives

530 of both interlocutors and trades off between them. In principle, the model of Heller and

531 colleagues (2016) and the *integration model* could easily be combined with one another.

532 However, because common ground in our study was established by a shared interaction and

533 not by the online convergence of different perspectives, it would require a different

534 experimental design to test the combined model.

535   The primary source of developmental change in our model is age related changes in the

536 propensity to make the individual inferences. As they get older, children expect speakers to

537 be more informative and to be more likely to follow common ground, but the process by

538 which the two information sources are integrated at any given age is assumed to be the same.

539 Other developmental models are also worth exploring in future work; one possible candidate

540 would be a model in which the integration process itself changes with age.

541   The developmental noise model reported for experiment 7 offers another way to

542 address the question of what changes with development. This model estimates a

543 developmental trajectory for the proportion of responses that are better explained by

544 random guessing than by the model structure. If such a model would find that model fit is

545 comparable for younger and older children but that the noise parameter through which this

546 fit is achieved decreases with age, we might conclude that cognitive abilities that have to do

547 with task demands are the major locus of change rather than abilities that have to do with

548 integrating information. In the developmental noise model in experiment 7, we found that

549 noise decreased with age but, at the same time, that the resulting model fit was substantially

550 worse for younger children. However, rather than a difference in how information is

551 integrated, we think that a lack of variation in children's responses is the reason for this poor

552 model fit. The strongest evidence for developmental changes in integration would come in a

553 case where younger children showed evidence of above/below-chance judgment in the

554 combined task that was distinct from that predicted by the two above/below-chance

555 component tasks. Such a comparison would require more precision (either via more trials or

556 more participants) than our current experiment affords, however.

557   Studying how multiple types of pragmatic cues are balanced contributes to a more

558 comprehensive understanding of word learning. In the current study, participants inferred

559 the referent by integrating non-linguistic cues (speakers pointing to a table) with

560 assumptions about speaker informativeness and common ground information, going beyond

561 previous experimental work in measuring how these information sources were combined. The

562 real learning environment is far richer than what we captured in our experimental design,

563 however. For example, in addition to multiple layers of social information, children can rely

564 on semantic and syntactic features of the utterances as cues to meaning (Clark, 1973;

565 Gleitman, 1990). Across development, children learn to recruit these different sources of

information and integrate them. RSA models allow for the inclusion of semantic information

as part of the utterance (Bergen et al., 2016) and it will be a fruitful avenue for future

research to model the integration of linguistic and pragmatic information across development.

To conclude, our work here shows how computational models of language comprehension can

be used as powerful tools to explicate and test hypotheses about information integration

across development.

[572]   **References**

[573]   Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early

[574]   word learning. *Child Development*, *67*(2), 635–645.

[575]   Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar

[576]   alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

[577]   Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models

[578]   using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

[579]   Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the

[580]   communicative intentions behind gestures in a hiding game. *Developmental Science*,

[581]   *8*(6), 492–499.

[582]   Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic

[583]   inference. *Semantics and Pragmatics*, *9*.

[584]   Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language.

[585]   *Annual Review of Developmental Psychology*, *1*(1), 223–249.

[586]   Bohn, M., & Koymen, B. (2018). Common ground and development. *Child Development*

[587]   *Perspectives*, *12*(2), 104–108.

[588]   Bohn, M., Le, K., Peloquin, B., Koymen, B., & Frank, M. C. (2020). Children's

[589]   interpretation of ambiguous pronouns based on prior discourse. *PsyArXiv*.

[590]   doi:10.31234/osf.io/gkhez

[591]   Bohn, M., Zimmermann, L., Call, J., & Tomasello, M. (2018). The social-cognitive basis of

[592]   infants' reference to absent entities. *Cognition*, *177*, 41–48.

[593]   Braginsky, M., Tessler, M. H., & Hawkins, R. (2019). *Rwebppl: R interface to webppl.*

Retrieved from https://github.com/mhtess/rwebppl

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190.

Bruner, J. (1983). *Child's talk: Learning to use language*. New York: Norton.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 65–110). New York: Academic Press.

Clark, E. V. (2009). *First language acquisition*. Cambridge: Cambridge University Press.

Clark, E. V. (2015). Common ground. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (Vol. 87, pp. 328–353). John Wiley & Sons.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cohn-Gordon, R., Goodman, N., & Potts, C. (2018). Pragmatically informative image captioning with character-level inference. *arXiv Preprint arXiv:1804.05417*.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of samuelson and smith. *Developmental Science*, *7*(1), 33–41.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games.

616   *Science*, *336*(6084), 998–998.

617   Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that

618   speakers are informative. *Cognitive Psychology*, *75*, 80–96.

619   Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential

620   intentions to model early cross-situational word learning. *Psychological Science*,

621   *20*(5), 578–585.

622   Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using

623   tablets to collect data from young children. *Journal of Cognition and Development*,

624   *17*(1), 1–17.

625   Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify

626   ambiguous requests. *Child Development*, *78*(2), 493–502.

627   Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1),

628   3–55.

629   Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic

630   inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

631   Goodman, N. D., & Stuhlmüller, A. (2014). The design and implementation of probabilistic

632   programming languages. http://dippl.org.

633   Graham, S. A., San Juan, V., & Khu, M. (2017). Words are not enough: How preschoolers'

634   integration of perspective and emotion informs their referential understanding.

635   *Journal of Child Language*, *44*(3), 500–526.

636   Grice, H. P. (1991). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

637   Grosse, G., Moll, H., & Tomasello, M. (2010). 21-month-olds understand the cooperative

logic of requests. *Journal of Pragmatics*, *42*(12), 3377–3383.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *89*(6), e572–e593.

Khu, M., Chambers, C. G., & Graham, S. A. (2020). Preschoolers flexibly shift between speakers' perspectives during real-time language comprehension. *Child Development*, *91*(3), e619–e634.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge, MA: MIT press.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(3), 403–422.

Monroe, W., Hawkins, R. X., Goodman, N. D., & Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, *5*, 325–338.

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148–165.

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science*, *42*, 974–1008.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Nilsen, E. S., Graham, S. A., & Pettigrew, T. (2009). Preschoolers' word mapping: The interplay between labelling context and specificity of speaker information. *Journal of Child Language*, *36*(3), 673–684.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

O'Neill, D. K., & Topolovec, J. C. (2001). Two-year-old children's sensitivity to the referential (in) efficacy of their own pointing gestures. *Journal of Child Language*, *28*(1), 1–28.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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*(4), 859–864.

Saylor, M. M., Sabbagh, M. A., Fortuna, A., & Troseth, G. (2009). Preschoolers use speakers' preferences to learn words. *Cognitive Development*, *24*(2), 125–132.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18.

Sperber, D., & Wilson, D. (2001). *Relevance: Communication and cognition* (2nd ed.). Cambridge, MA: Blackwell Publishers.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development, 11*(2), 176–190.

Sullivan, J., Boucher, J., Kiefer, R. J., Williams, K., & Barner, D. (2019). Discourse coherence as a cue to reference in word learning: Evidence for discourse bootstrapping. *Cognitive Science, 43*(1), e12702.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT press.

Tomasello, M. (2009). *Constructing a language.* Cambridge, MA: Harvard University Press.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences, 109*(32), 12933–12937.

Wang, S., Liang, P., & Manning, C. D. (2016). Learning language games through interaction. In *54th annual meeting of the association for computational linguistics, acl 2016* (pp. 2368–2378). Association for Computational Linguistics (ACL).

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review, 114*(2), 245.

Yoon, E. J., & Frank, M. C. (2019). The role of salience in young children's processing of ad hoc implicatures. *Journal of Experimental Child Psychology, 186*, 99–116.

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*(2), 635–645.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar
    alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

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

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the
    communicative intentions behind gestures in a hiding game. *Developmental Science*,
    8(6), 492–499.

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic
    inference. *Semantics and Pragmatics*, *9*.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language.
    Annual Review of Developmental Psychology, *1*(1), 223–249.

Bohn, M., & Koymen, B. (2018). Common ground and development. *Child Development
    Perspectives*, *12*(2), 104–108.

Bohn, M., Le, K., Peloquin, B., Koymen, B., & Frank, M. C. (2020). Children's
    interpretation of ambiguous pronouns based on prior discourse. *PsyArXiv*.
    doi:10.31234/osf.io/gkhez

Bohn, M., Zimmermann, L., Call, J., & Tomasello, M. (2018). The social-cognitive basis of
    infants' reference to absent entities. *Cognition*, *177*, 41–48.

Braginsky, M., Tessler, M. H., & Hawkins, R. (2019). *Rwebppl: R interface to webppl*.
    Retrieved from https://github.com/mhtess/rwebppl

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential
    precedents during interactive dialog. *Journal of Memory and Language*, *61*(2),
    171–190.

Bruner, J. (1983). *Child's talk: Learning to use language.* New York: Norton.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first
language. In T. Moore (Ed.), *Cognitive development and acquisition of language* (pp.
65–110). New York: Academic Press.

Clark, E. V. (2009). *First language acquisition.* Cambridge: Cambridge University Press.

Clark, E. V. (2015). Common ground. In B. MacWhinney & W. O'Grady (Eds.), *The
handbook of language emergence* (Vol. 87, pp. 328–353). John Wiley & Sons.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Cohn-Gordon, R., Goodman, N., & Potts, C. (2018). Pragmatically informative image
captioning with character-level inference. *arXiv Preprint arXiv:1804.05417.*

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity
to speakers' intent: An alternative account of samuelson and smith. *Developmental
Science, 7*(1), 33–41.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a
statistically optimal fashion. *Nature, 415*(6870), 429.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of
cross-situational word learning. *Cognitive Science, 34*(6), 1017–1063.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games.
Science, 336(6084), 998–998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that
speakers are informative. *Cognitive Psychology, 75*, 80–96.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential

intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17.

Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests. *Child Development*, *78*(2), 493–502.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

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

Goodman, N. D., & Stuhlmüller, A. (2014). The design and implementation of probabilistic programming languages. http://dippl.org.

Graham, S. A., San Juan, V., & Khu, M. (2017). Words are not enough: How preschoolers' integration of perspective and emotion informs their referential understanding. *Journal of Child Language*, *44*(3), 500–526.

Grice, H. P. (1991). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Grosse, G., Moll, H., & Tomasello, M. (2010). 21-month-olds understand the cooperative logic of requests. *Journal of Pragmatics*, *42*(12), 3377–3383.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the
        probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers:
        Exploring children's deficits in scalar implicature. *Child Development*, *89*(6),
        e572–e593.

Khu, M., Chambers, C. G., & Graham, S. A. (2020). Preschoolers flexibly shift between
        speakers' perspectives during real-time language comprehension. *Child Development*,
        91(3), e619–e634.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational*
        implicature. Cambridge, MA: MIT press.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual
        availability and prior discourse on young children's use of referring expressions.
        Applied Psycholinguistics, *27*(3), 403–422.

Monroe, W., Hawkins, R. X., Goodman, N. D., & Potts, C. (2017). Colors in context: A
        pragmatic neural model for grounded language understanding. *Transactions of the*
        Association for Computational Linguistics, *5*, 325–338.

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared
        knowledge about object identity in real-time referential processing. *Cognition*, *142*,
        148–165.

Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the
        probabilistic combination of multiple perspectives. *Cognitive Science*, *42*, 974–1008.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's
        on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Nilsen, E. S., Graham, S. A., & Pettigrew, T. (2009). Preschoolers' word mapping: The interplay between labelling context and specificity of speaker information. *Journal of Child Language*, *36*(3), 673–684.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

O'Neill, D. K., & Topolovec, J. C. (2001). Two-year-old children's sensitivity to the referential (in) efficacy of their own pointing gestures. *Journal of Child Language*, *28*(1), 1–28.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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*(4), 859–864.

Saylor, M. M., Sabbagh, M. A., Fortuna, A., & Troseth, G. (2009). Preschoolers use speakers' preferences to learn words. *Cognitive Development*, *24*(2), 125–132.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18.

Sperber, D., & Wilson, D. (2001). *Relevance: Communication and cognition* (2nd ed.). Cambridge, MA: Blackwell Publishers.

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*(2), 176–190.

Sullivan, J., Boucher, J., Kiefer, R. J., Williams, K., & Barner, D. (2019). Discourse coherence as a cue to reference in word learning: Evidence for discourse

819   bootstrapping. *Cognitive Science*, *43*(1), e12702.

820 Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995).

821   Integration of visual and linguistic information in spoken language comprehension.

822   *Science*, *268*(5217), 1632–1634.

823 Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT press.

824 Tomasello, M. (2009). *Constructing a language.* Cambridge, MA: Harvard University Press.

825 Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize

826   that speech can communicate unobservable intentions. *Proceedings of the National*

827   *Academy of Sciences*, *109*(32), 12933–12937.

828 Wang, S., Liang, P., & Manning, C. D. (2016). Learning language games through interaction.

829   In *54th annual meeting of the association for computational linguistics, acl 2016* (pp.

830   2368–2378). Association for Computational Linguistics (ACL).

831 Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological*

832   *Review*, *114*(2), 245.

833 Yoon, E. J., & Frank, M. C. (2019). The role of salience in young children's processing of ad

834   hoc implicatures. *Journal of Experimental Child Psychology*, *186*, 99–116.

Declarations of interest

None.

Author Contributions

M. Bohn and M.C. Frank conceptualized the study, M. Merrick collected the data, M. Bohn and M.H. Tessler analyzed the data, M. Bohn, M. H. Tessler and M.C. Frank wrote the manuscript, all authors approved the final version of the manuscript.

Acknowledgments

Funding