

¹ Predicting information integration in pragmatic word learning across development

² Manuel Bohn^{1,2}, Michael Henry Tessler³, & Michael C. Frank¹

³ ¹ Department of Psychology, Stanford University

⁴ ² Leipzig Research Center for Early Child Development, Leipzig University

⁵ ³ Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

⁶ Author Note

⁷ MB received funding from the European Union's Horizon 2020 research and innovation
⁸ programme under the Marie Skłodowska-Curie grant agreement no. 749229. We thank
⁹ Jacqueline Quirke, Sabina Zacco and especially Megan Merrick for help with the data
¹⁰ collection.

¹¹ Correspondence concerning this article should be addressed to Manuel Bohn,
¹² Department of Psychology, 450 Serra Mall, Stanford, CA 94305. E-mail: bohn@stanford.edu

14

Abstract

15 Language is used and learned in complex social settings. To infer what a speaker means,
16 listeners have to go beyond the literal meaning of words and rely on pragmatic inferences.
17 During language acquisition, the same processes can help children to learn the meaning of
18 novel words. Yet, pragmatic inferences require balancing multiple sources of social
19 information. Here we study this integration process. We isolate two types of pragmatic
20 inference and formalize how they should be integrated in a Bayesian pragmatics model. We
21 present a series of experiments with preschool children and adults that suggest that
22 information integration is best described as a form of probabilistic inference.

23 *Keywords:* language acquisition, social cognition, pragmatics, Bayesian modeling

24 Word count: X

25 Predicting information integration in pragmatic word learning across development

26 **Introduction**

27 Language use and language learning require integrating and balancing different sources
28 of information. And even though linguistic cues, such as conventional syntax and semantics,
29 may provide strong evidence about the intended meaning of an utterance, additional
30 pragmatic inferences are necessary arrive at a precise interpretation (Levinson, 2000; Sperber
31 & Wilson, 2001). A second route to meaning is even more important in the process of
32 language learning. In this case, inferences based on the social embedding of the utterance are
33 key to understanding its meaning and learning the words contained therein (E. V. Clark,
34 2009; Tomasello, 2009). In this paper, we explore how different forms of social information
35 are integrated during word learning across development. More specifically, we look at
36 information integration as a form of probabilistic inference.

37 Pragmatic inferences rest on assumptions about how communication *should* proceed.
38 Interlocutors mutually expect each other to communicate in a cooperative way and this
39 mutual expectation serves as the basis for inferences that go beyond the literal meaning of
40 an utterance (Grice, 1991; Sperber & Wilson, 2001; Tomasello, 2008). Paul Grice (1991)
41 famously summarized this notion in his *Cooperative Principle*, which reads: “Make your
42 contribution such as is required, at the stage at which it occurs, by the accepted purpose or
43 direction of the talk exchange in which you are engaged.” Two sets of expectations follow
44 from this principle. On the one hand, there is the general expectation that one follows the
45 principle and communicates in a relevant and informative way. On the other hand, there is
46 the expectation that communication is further specific to the communicative partner, that is,
47 that it is adjusted to the common ground shared between speaker and listener (Bohn &
48 Koymen, 2018; H. H. Clark, 1996).

49 Importantly, these expectations can also support language development (E. V. Clark,
50 2009; Tomasello, 2009). Starting in infancy, children generally expect adults’ communicative
51 acts to provide them with relevant information (e.g Behne, Carpenter, & Tomasello, 2005).

52 And preschool children learn novel words by assuming that speakers are informative (Frank
53 & Goodman, 2014). In the absence of any additional information about the potential
54 referent objects or the speaker, children expected a novel word to be intended to refer to an
55 object it describes in a unique way (see also Stiller, Goodman, & Frank, 2015). Person
56 specific common ground information supports communication in infancy (e.g. Bohn,
57 Zimmermann, Call, & Tomasello, 2018; Saylor, Ganea, & Vázquez, 2011) and facilitates
58 word learning in slightly older children. For example, two-year-olds keep track of which
59 object is new to a speaker to decide which object the speaker is referring to when they use a
60 novel word (Akhtar, Carpenter, & Tomasello, 1996; Diesendruck, Markson, Akhtar, &
61 Reudor, 2004). Furthermore, when three to five year old children learn about a speaker's
62 preference, they later use this information to identify the meaning of an unknown word
63 coming from the same speaker (Saylor, Sabbagh, Fortuna, & Troseth, 2009). These studies
64 suggest early emerging pragmatic competences. Yet, communicative interactions in
65 naturalistic contexts have multiple layers and require balancing expectations (H. H. Clark,
66 1996). So how are different expectations integrated?

67 The Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman &
68 Frank, 2016) offers a formal framework for addressing this integration problem. RSA models
69 are characterized by their recursive structure. Each agent in the recursion is modeled as a
70 Bayesian reasoner; thus, information integration is treated as a process of probabilistic
71 inference. The expectation that speakers communicate informatively is already encoded in
72 the structure of the model: Speakers choose utterances that help listeners in disambiguating
73 referents. Common ground information can be integrated in the model as a shared prior
74 probability of referents in the context of the utterance. Thus, a natural locus for information
75 integration within RSA models is the trade off between the prior probability of a referent
76 and the likelihood of that referent given the current utterance.

77 Here we apply this rational, pragmatic account of communication to the study of
78 information integration during word learning. We do so from a developmental perspective

79 and include children, continuously sampled between three and five years of age, as well as
80 adults. For both groups, we first test speaker-specific and common ground expectations
81 independently. Based on this data, we generate model predictions about how expectations
82 should be integrated. Finally, we compare these predictions to newly collected data and use
83 model comparisons to test different hypothesis about how expectations are integrated.

84

Adults

85 **Assessment of general and speaker specific expectations (Experiment 1 and 2)**

86 As a first step, we assessed whether adults make general and common ground
87 inferences when tested separately. In experiment 1, participants could infer which object a
88 novel word referred to by assuming that the speaker communicated in an informative way.
89 The speaker was located between two tables, one with two novel objects, A and B, and the
90 other with only object A (Fig. 1A). When the speaker turned to the table with the two
91 objects (A and B) and used an novel word to request one of them, participants could infer
92 that the word referred to object B. This follows from the counter-factual inferences that, if
93 the (informative) speaker would have wanted to refer to object A, they would have turned to
94 the table with the single object, because this would have been the least ambiguous way to
95 refer to this object. As a consequence, turning to the table with two objects most likely
96 reflects an intention to refer to object B. In the control condition, both tables contained both
97 objects and no inference could be made based on the speaker's behavior.

98 Participants made the inference consistent with the assumption that the speaker
99 communicated informatively: They selected object B above chance in the test condition
100 ($t(39) = 5.51, p < .001$, see Fig. 2) and more often compared to the control condition ($\beta =$
101 $1.49, se = 0.50, p = .003$).

102 In experiment 2, we tested if participants use common ground information that is
103 specific to a speaker to infer the referent of a novel word. In the preference manipulation
104 (Fig. 1B . left), the speaker expressed preference for one of two objects. When the speaker

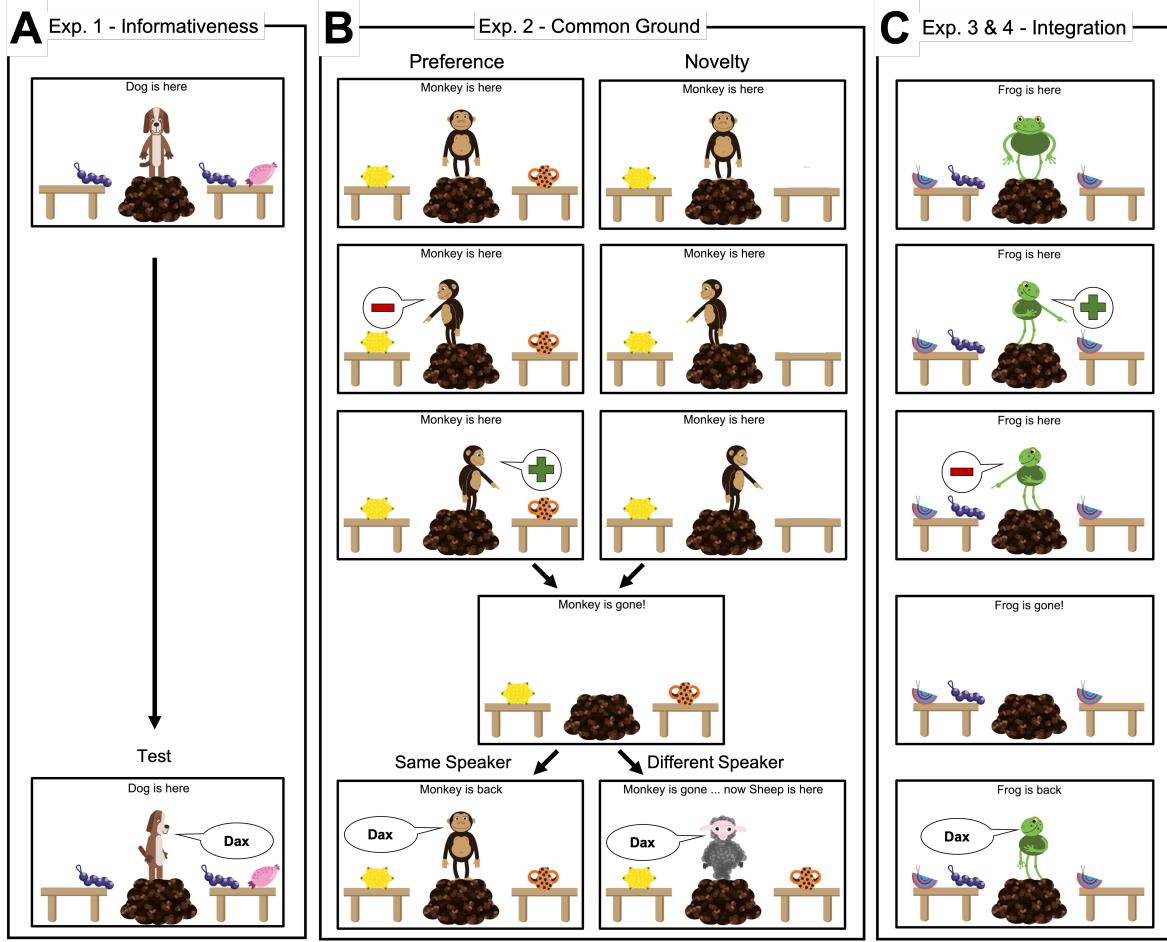


Figure 1. Schematic experimental procedure with screenshots from the adult experiments. In all conditions, at test (bottom), the speaker ambiguously requested an object using a non-word (e.g. “dax”). Participants clicked on the object they thought the speaker referred to. Informativeness (Experiment 1, left) translated to making one object less frequent in context. Common ground (Experiment 2, middle) was manipulated by making one object preferred by or new to the speaker. Green plus signs represent utterances that expressed preference and red minus of dispreference (see main text for details). Experiment 3 (right) combined manipulations. When expressing e.g. preference for an object on a table with two objects (panel 3 from top), the respective object was temporarily enlarged. Condition for Experiment 3 shown here: preference - same speaker - incongruent.

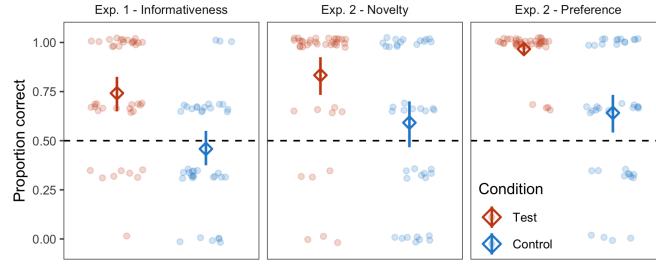


Figure 2. Results from experiment 1 and 2 for adults. For preference and novelty, control refers to a different speaker (see Fig. 1B). Transparent dots show data from individual participants, diamonds represent condition means, error bars are 95% CIs. Dashed line indicates performance expected by chance.

105 later used a novel word to requested an object, participants could infer the referent by
 106 assuming that the speaker was talking about the object they liked. In a control condition, a
 107 different speaker, whose preferences were unknown, made the request. Adults selected the
 108 preferred object above chance when the same speaker made the request ($t(39) = 29.14, p <$
 109 .001, see Fig. 2) and more so compared to when a different speaker asked for an object ($\beta =$
 110 2.97, $se = 0.70, p < .001$).

111 In the novelty manipulation (Fig 1B - right), the speaker encountered one object on
 112 one of the tables. Later, a second object appeared. The second object was therefore new to
 113 the speaker. When the same speaker later used a novel word to requested an object (in a
 114 slightly excited way), participants could infer that it was the new object they are referring to.
 115 In contrast, when a different speaker, to whom both objects were equally new, made a
 116 request, no inference could be made based on the speaker's state of knowledge. In line with
 117 this reasoning, participants selected the new object above chance with the same speaker
 118 ($t(39) = 6.77, p < .001$, see Fig. 2), and more often compared to the different speaker ($\beta =$
 119 1.70, $se = 0.39, p < .001$). Taken together, experiment 1 and 2 showed that adults use
 120 general and speaker specific expectations to identify the referent of a novel word. In the
 121 following, we studied how these expectations are integrated when combined experimentally.

¹²² **Model predictions for information integration evaluated against new data**
¹²³ **(Experiment 3).**

¹²⁴ We modeled the integration of general and common ground expectations as a process
¹²⁵ of probabilistic inference. We used the results of experiment 1 and 2 to inform key
¹²⁶ parameters in the model’s architecture. Within the RSA framework, general expectations
¹²⁷ that speakers (or listeners) communicate in a cooperative and informative way are reflected
¹²⁸ in the model’s recursive structure. When interpreting an utterance, the hypothetical listener
¹²⁹ imagines that it was generated by a pragmatic speaker whose goal is to produce the most
¹³⁰ informative utterance. The pragmatic speaker is modeled as reasoning about what the most
¹³¹ informative utterance would be for a naive listener, who does not know the novel words.
¹³² Importantly, the utterance in this case includes not just the novel word but also the turning
¹³³ to one of the tables. The conditional probability that the listener interprets an utterance as
¹³⁴ referring to a particular referent is defined as:

$$P_L(r_s|u) \propto P_S(u|r_s)P_S(r_s)$$

¹³⁵ Here, the term $P_S(u|r_s)$ denotes the likelihood that the pragmatic speaker will produce
¹³⁶ a particular utterance u to refer to a referent r . It is defined in terms of a utility function
¹³⁷ $U_S(u; s)$ consisting of the surprisal of u for a naive listener, who interprets u according its
¹³⁸ literal semantics:

$$P_S(u|r_s) \propto \exp(\alpha U_S(u; s))$$

¹³⁹ This utility is maximized in order to decide which utterance to use to communicate
¹⁴⁰ about a particular referent, reflecting the expectation that the speaker communicates in an
¹⁴¹ informative way. The absolute strength of $P_S(u|r_s)$ depends on a scalar value, α , which can
¹⁴² be interpreted as an indicator of how rational the speaker is in choosing utterances (i.e. as α
¹⁴³ increases, the speaker is more likely to choose the most informative utterance). We used the

¹⁴⁴ data from experiment 1 to specify how rational participants thought speakers were in the
¹⁴⁵ current setup. That is, we inferred which value of α would generate model predictions
¹⁴⁶ (assuming equal prior probability for each object) that corresponded to the average
¹⁴⁷ proportion of correct responses measured in experiment 1.

¹⁴⁸ We treated common ground expectations as changes in the prior probability that a
¹⁴⁹ particular referent is being referred to ($P_S(r_s)$ in the first equation). For example, if the
¹⁵⁰ speaker expresses preference for object A, and the same speaker later requests an object,
¹⁵¹ then A has a higher probability of being the referent of a subsequent utterance, regardless of
¹⁵² any additional pragmatic considerations. We used the proportions measured in experiment 2
¹⁵³ to specify the probability distribution across referents for the different common ground
¹⁵⁴ manipulations (novelty or preference, same or different speaker).

¹⁵⁵ With α and $P_S(r_s)$ fixed based on experiment 1 and 2, we generated model predictions
¹⁵⁶ for situations in which general and common ground expectations had to be integrated (Fig.
¹⁵⁷ 1C). Expectations could either be congruent or incongruent. If, for example, object A was
¹⁵⁸ the more informative object (unique to that table) and also the one the speaker expressed
¹⁵⁹ preference for, expectations were congruent. Next, the speaker who made the request could
¹⁶⁰ either be the same as the one in the previous interaction or a different one. Finally, common
¹⁶¹ ground information could be manipulated in the form of preference or novelty. This resulted
¹⁶² in a total of 2 (novelty or preference) x 2 (same or different speaker) x 2 (congruent or
¹⁶³ incongruent) = 8 conditions for which we generated model predictions. In order to capture
¹⁶⁴ that behavioral data is to some extend noisy, all models included a noise parameter. Noise
¹⁶⁵ parameters ranged between 0 and 1 and reflect the proportion of responses that are
¹⁶⁶ estimated (based on the data) to be random instead of in line with model predictions.
¹⁶⁷ Including the noise parameter greatly improved the model fit to the data (see SI Appendix
¹⁶⁸ for details). We did not preregister the inclusion of a noise parameter for experiment 3 but
¹⁶⁹ did so for all subsequent experiments.

¹⁷⁰ In experiment 3, we collected new behavioral data for each of the eight conditions.

171 Experimentally, we combined the procedures of experiment 1 and 2. The test setup was
172 identical to experiment 1, however, before making a request, the speaker interacted with the
173 objects so that some of them were preferred by or new to them (Fig. 1C). We discuss and
174 visualize the results as the proportion with which participants chose the more informative
175 object (i.e., the object that would be the more informative referent when only general
176 expectations are considered). Before comparing responses to model predictions we fit a
177 generalized linear mixed model (GLMM) to the data to test if participant's responses varied
178 across conditions. Participants distinguished between congruent and incongruent trials when
179 the speaker remained the same (model term: `alignment x speaker`; $\beta = -2.90$, $se = 0.56$, p
180 $< .001$).

181 Model predictions were highly correlated with the average response in each condition
182 (Fig. 3A and B). We compared the fit of the pragmatics model described above to two
183 alternative models: the *flat prior* model ignored the speaker specific expectation and the
184 *prior only* model ignored the informativeness inference. This analysis tests whether the
185 trade-off between expectations in the pragmatics model captured the structure in the data
186 better compared to models that assume that participants only consider one type of
187 expectation at a time. Model fit was much better for the pragmatic compared to the flat
188 prior (Bayes Factor (BF) = 4.2e+53) or the prior only model (BF = 2.5e+34), suggesting
189 that participants considered and integrated both types of expectations. The estimated
190 proportion of random responses according to the pragmatics model was 0.30 (95% Highest
191 Density Interval (HDI): 0.23 - 0.36). This value was considerably lower for the pragmatics
192 model, compared to the alternative models (see SI Appendix), lending additional support to
193 the conclusion that the pragmatics model better captured the behavioral data.

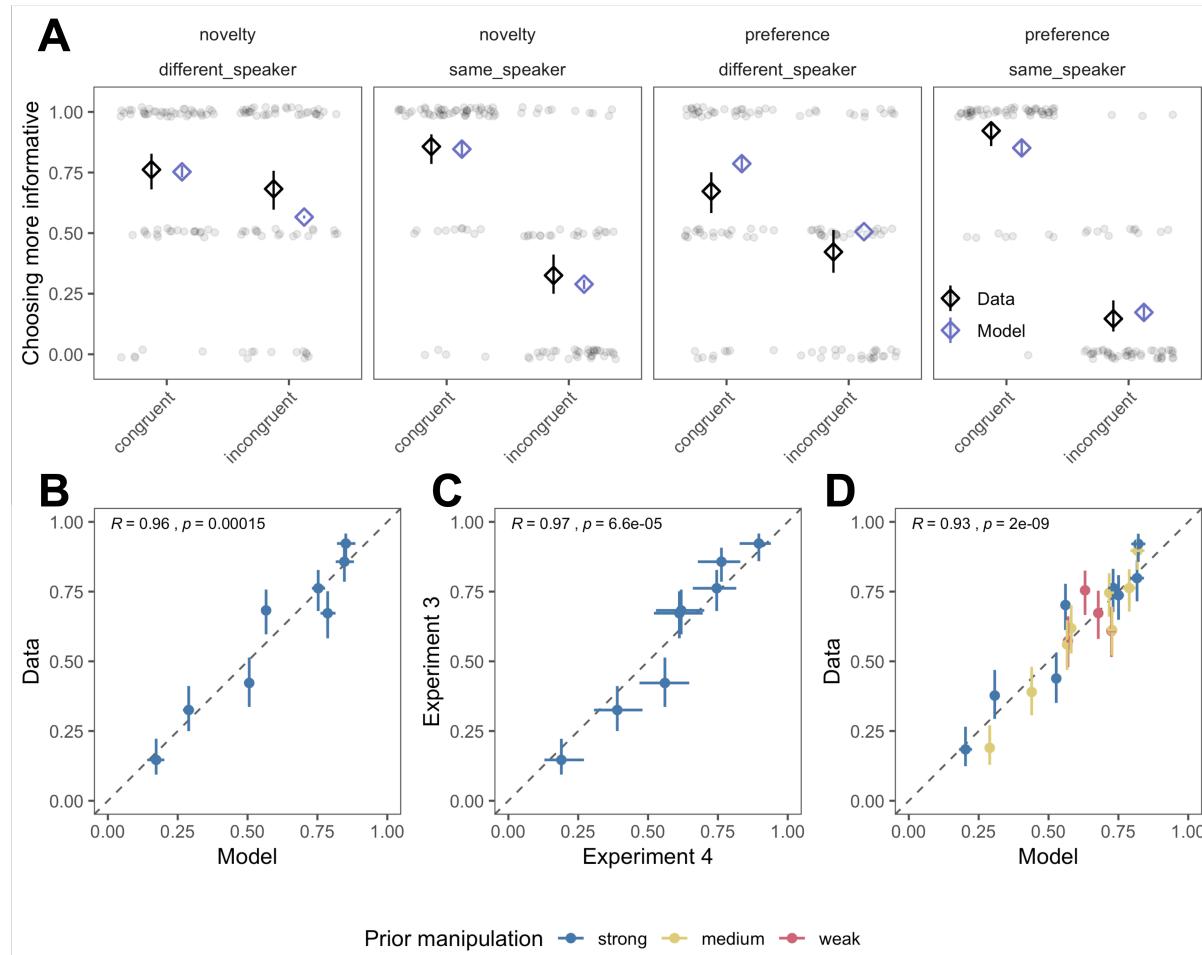


Figure 3. Results from experiment 3 and 4 for adults. Data and model predictions by condition for experiment 3 (A). Transparent dots show data from individual participants, diamonds represent condition means. Correlation between model predictions and data in experiment 3 (B), between data in experiment 3 and data for the strong prior manipulation in experiment 4 (direct replication - 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.

¹⁹⁴ **Replication and extension to different levels of speaker specific expectation**

¹⁹⁵ **(Experiment 4).**

¹⁹⁶ The goal of experiment 4 was twofold: First, to replicate the results of experiment 3
¹⁹⁷ and second, to test the generalizability of the pragmatics model to a larger range of

198 conditions. Compared to experiment 3, we manipulated the strength of the common ground
 199 expectations. For both, preference and novelty, we planned to have a strong, a medium and
 200 a weak manipulation. Expectation strength was manipulated by changing the way the
 201 speaker interacted with the objects prior to the request and measured as the proportion with
 202 which participants chose the preferred/novel object in the same speaker condition (see SI
 203 Appendix for results). For novelty, we succeeded in finding three qualitatively different
 204 manipulations. For preference, we piloted a number of manipulations but did not find a way
 205 to indicate only weak preference. Experiment 4 therefore included strong and medium
 206 manipulations for novelty and preference and a weak manipulation for novelty. The general
 207 expectation was the same as in experiment 3. For each level we crossed general and common
 208 ground expectations in the same way as in experiment 3, resulting in a total of 8 (strong) +
 209 8 (medium) + 4 (weak) = 20 conditions. The strong manipulation was a direct replication of
 210 experiment 3. Model predictions were generated in the same way as in experiment 3.

211 Again, before evaluating model predictions, we fit GLMMs to the data for each level of
 212 common ground manipulation. Participants distinguished between congruent and
 213 incongruent trials when the speaker remained the same for the strong ($\beta = -1.60$, $se = 0.52$,
 214 $p = .002$)¹ and medium ($\beta = -2$, $se = 0.51$, $p < .001$), but not the weak manipulation ($\beta =$
 215 -0.68 , $se = 0.55$, $p = .219$, model term: `alignment x speaker` in all cases). This pattern
 216 validates our manipulations. If the speaker specific manipulation is weak, changing the
 217 speaker has no differential impact on the informativeness inference. Results from the strong
 218 manipulation in experiment 4 further replicated findings from experiment 3 (Fig. 3C).

219 Model predictions from the pragmatics model were again highly correlated with the
 220 average response in each condition (Fig. 3D). We evaluated model fit for the same models as
 221 in experiment 3 and found again that the pragmatics model fit the data much better
 222 compared to the flat prior ($BF = 4.7e+71$) or the prior only model ($BF = 8.9e+82$). The

¹Here, this effect differed between novelty and preference: model term manipulation x alignment x speaker;
 $\beta = -2.96$, $se = 0.73$, $p < .001$

223 inferred level of noise based on the data for the pragmatics model was 0.36 (95% HDI: 0.31 -
224 0.41), which was again lower compared to the alternative models (see SI Appendix).

225 Summarizing experiment 3 and 4, we saw that adults flexibly integrated different
226 expectations about a speaker in a word learning scenario. This process was best described by
227 a model that treated information integration as a form of probabilistic inference. Next, we
228 took a developmental perspective and studied if the same model could be used to describe
229 children's word learning.

230 **Children**

231 **Assessment of general and speaker specific expectations (Experiment 1 and 2)**

232 We approached children's word learning in the same way as adults'. First, we measured
233 the strength of general (informativeness) and common ground (preference) expectations and
234 used these measurements to set model parameters. Then we generated model predictions for
235 how expectations should be integrated. Finally, we compared model predictions to new
236 behavioral data. In this, we took a truly developmental perspective. We sampled children
237 continuously across the entire age range and generated model predictions for the average
238 developmental trajectory in each condition. In all the following experiments, we tested
239 children between 3.0 and 4.9 years of age. For experiment 1 and 2, we also tested
240 two-year-olds but did not find sufficient evidence that they make general and/or common
241 ground inferences in the tasks we used.

242 Children were only tested in the test condition in experiment 1. To compare children's
243 performance to chance level, we binned age by year. Four-year-olds selected the more
244 informative object (i.e. the object that was unique to the location the speaker turned to)
245 above chance ($t(29) = 2.80, p = .009$). Three-year-olds, on the other hand, did not ($t(31) =$
246 $-1.31, p = .198$). Consequently, when we fit a GLMM to the data with age as a continuous
247 predictor, performance increased with age ($\beta = 0.38, se = 0.11, p < .001$, see Fig. 4). Thus,
248 children's expectation that speakers communicate in an informative way increased with age.

249 In experiment 2, we assessed whether children use common ground information to
 250 identify the referent of a novel word. We tested children with the novelty as well as the
 251 preference manipulation but found little evidence that children distinguished between
 252 requests made by the same speaker or a different speaker in the case of novelty. We therefore
 253 dropped this manipulation and focused on preference for the remainder of the study.

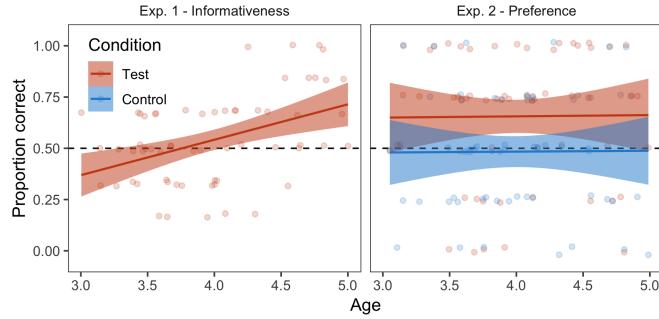


Figure 4. Results from experiment 1 and 2 for children. For preference, control refers to the different speaker condition (see Fig. 1B). Transparent dots show data from individual participants, regression lines show fitted linear models with 95% CIs. Dashed line indicates performance expected by chance.

254 For preference, four-year-olds selected the preferred object above chance when the same
 255 speaker made the request ($t(30) = 4.14, p < .001$), whereas three-year-olds did not ($t(29) =$
 256 $1.62, p = .117$). However, when we fit a GLMM to the data with age as a continuous
 257 predictor, we found an effect of speaker identity ($\beta = 0.92, se = 0.30, p = .002$) but no effect
 258 of age ($\beta = 0.04, se = 0.18, p = .83$) or interaction between speaker identity and age ($\beta =$
 259 $-0.04, se = 0.24, p = .886$, Fig. 4). Thus, children across the age range used common ground
 260 information to infer the referent of a novel word.

261 **Model predictions for information integration evaluated against new data**
 262 **(Experiment 3)**

263 We used the measurements from experiment 1 and 2 to specify α and $P_S(r_s)$ in the
 264 pragmatics model. However, instead of setting these parameters to a single value, we used

the data to infer the intercept and slope that best described the developmental trajectory for each parameter. For $P_S(r_s)$, this was done separately for the same speaker and the different speaker condition (see SI Appendix for details). These parameter settings were then used to generate age sensitive model predictions in 2 (same or different speaker) x 2 (congruent or incongruent) = 4 conditions. As for adults, all models included a noise parameter, which was estimated based on the data.

In experiment 3, we combined the procedures of experiment 1 and 2 and collected new data from children between 3.0 and 4.9 years of age in each of the four conditions (Fig. 1C). Before comparing the data to model predictions, we fit a GLMM to the data to test if children distinguished between conditions. Children's propensity to differentiate between congruent and incongruent trials for the same or a different speaker increased with age (model term: `age x alignment x speaker`; $\beta = -1.18$, $se = 0.50$, $p = .018$).

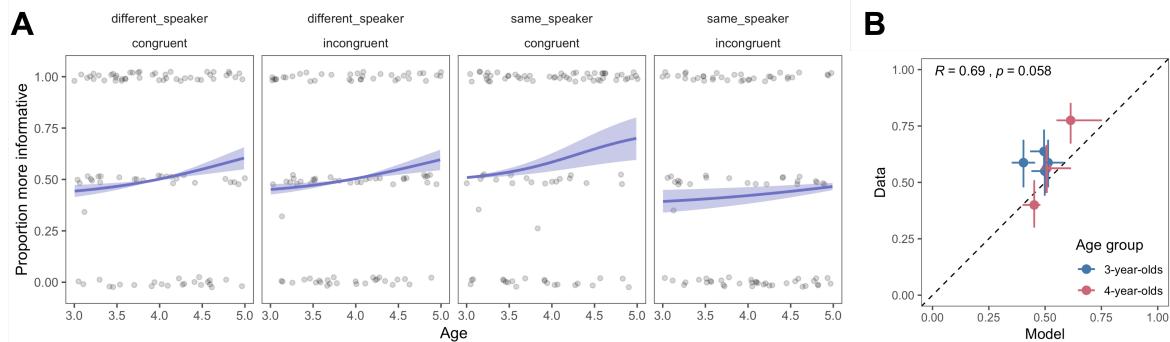


Figure 5. Results from experiment 3 for children. Purple lines show model predictions with 95% HDIs, transparent dots show data from individual participants (A). Correlation between model predictions and condition means binned by year (B). Coefficient and p-value are based on Pearson correlation statistic. Error bars represent 95% HDIs.

Next, we evaluated the model predictions (Fig 5A). For the correlational analysis, we binned model predictions and data by year. There was a substantial correlation between the predicted and measured average response in each age bin (Fig. 5B). For the model comparison, we treated age continuously. As for adults, we found a much better model fit for

the pragmatics model compared to the flat prior ($BF = 577$) or the prior only model ($BF = 10560$). The inferred level of noise based on the data for the pragmatics model was 0.51 (95% HDI: 0.26 - 0.77), which was lower compared to the alternative models considered (see SI Appendix). The high level of inferred noise moved the model predictions in all conditions close to chance level (Fig. 5A). We therefore compared two additional sets of models with different parameterizations that emphasized differences between conditions in the model predictions more. This analysis was not preregistered. 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 pragmatics model provided a better fit compared to the alternative models (flat prior: parameter free $BF = 334$, developmental noise $BF = 16361$; prior only: parameter free $BF = 20$, developmental noise $BF = 1e+06$). Taken together, these results suggest that children's word learning is best described as a flexible integration of general and common ground expectations.

Discussion

Integrating different sources of information is an integral part of human communication. To infer the intended meaning of an utterance, listeners combine their knowledge of communicative conventions (semantics and syntax) with social expectations about their interlocutor. When learning new words, social information is especially vital. But how are different social expectations integrated during word learning? Here we explored the hypothesis that information integration is a process of probabilistic inference. We studied how expectations that speakers communicate in an informative way are contextualized by expectations that follow from prior interactions with the speaker. Across development, participants' learning was best predicted by a Bayesian pragmatics model in which expectations were flexibly traded-off between one another in a probabilistic way. Alternative models, which considered only one type of expectation, made worse predictions.

The pragmatics model goes beyond simply weighing expectations by some ratio. It

307 offers an explicit account about how integration may proceed: Because all objects are new
308 and their labels unknown, the initial probability of each object depends entirely on common
309 ground information. Now, the listener assumes the speaker to produce their utterance based
310 on the expected effect that their utterance will have *in light of* this shared common ground.
311 Thus, inferences are not computed separately but the result of an integrated process.

312 In the current model, common ground information was treated in the same way as
313 contextual salience (see e.g. Frank & Goodman, 2012). Interacting around an object
314 changed the probability that this object will be communicated about. In theory, similar
315 changes could be brought about by increasing the object's perceptual salience. Future work
316 could try to explicitly model the social-cognitive processes that gave rise to common ground
317 expectations and contrast social and perceptual salience.

318 We conceptualized developmental change as age related changes in the propensity to
319 make the individual inferences. That is, while the degree to which listeners expect speakers
320 to be informative or follow common ground changes with age, the process by which
321 expectations are integrated remains the same. This is a very parsimonious conception of
322 developmental change and others, in which the integration process itself changes with age,
323 are also plausible and worth exploring.

324 Studying how multiple types of social information are balanced contributes to a
325 comprehensive and ecologically valid model of word learning. In the current study,
326 participants inferred the referent by integrating non-linguistic cues (speakers turning and
327 looking to one table) with assumptions about speaker informativeness and common ground
328 information. However, the real learning environment is much richer than the what we
329 captured in our experimental design. In addition to multiple layers of social information,
330 children can rely on semantic and syntactic features of the utterances as cues to the meaning
331 of the words embedded therein (Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017;
332 E. V. Clark, 1973; Gleitman, 1990). Across development, children learn to recruit these
333 different sources of information and integrate them. Our studies show how computational

334 models of language use and comprehension can be used as powerful tools to explicate and
335 test hypothesis about how integration proceeds.

336 **Methods**

337 All experimental procedures, sample sizes and statistical analysis were pre-registered
338 (<https://osf.io/u7kxe/>). Experimental stimuli, data files and analysis scripts are freely
339 available in an online repository (<https://github.com/manuelbohn/mcc>).

340 **Participants**

341 Adult participants were recruited via Amazon Mechanical Turk (MTurk) and received
342 payment equivalent to an hourly wage of ~ \$9. Experiment 1 and each manipulation of
343 experiment 2 had $N = 40$ participants. Sample size in experiment 3 was $N = 121$. $N = 167$
344 participated in the experiments to measure the strong, medium and weak preference and
345 novelty manipulations. Finally, experiment 4 had $N = 286$ participants.

346 Children were recruited from the floor of the Children's Discovery Museum in San Jose,
347 California, USA. Parents gave informed consent and provided demographic information. We
348 collected data from a total of 243 children between 3.0 and 4.9 years of age. We excluded 15
349 due to less than 75% of reported exposure to English, five because they responded incorrect
350 on 2/2 training trials, three because of equipment malfunction and two because they quit
351 before half of the test trials were completed. The final sample size in each experiment was as
352 follows: $N = 62$ (41 girls, mean age = 4) in experiment 1, $N = 61$ (28 girls, mean age =
353 3.99) in experiment 2 and $N = 96$ (54 girls, mean age = 3.96) in experiment 3.

354 **Experimental procedures**

355 All experiments were framed as games in which participants would learn words from
356 animals. They were implemented in HTML/JavaScript as a website. Adults were directed to
357 the website via MTurk and responded by clicking objects. Children were guided through the
358 game by an experimenter and responded by touching objects on the screen. For each animal

359 character, we recorded a set of utterances (one native English speaker per animal) that were
360 used to provide information and make requests. All studies started with an introduction to
361 the animals and two training trials in which familiar objects were requested (car and ball).
362 Subsequent test trials in each condition were presented in a random order.

363 The setup of experiment 1 for adults is shown in Fig. 1A. In the beginning of each
364 trial, the animal introduced themselves (e.g. “Hi, I’m Dog”) and then turned towards the
365 table with the two objects. The same utterance was used to make a request in all adult
366 studies (“Oh cool, there is a [non-word] on the table, how neat, can you give me the
367 [non-word]?”). In the test condition, there was one object on the other table whereas in the
368 control condition there were two. Participants received six trials, three per condition.

369 The setup for experiment 2 is shown in Fig. 1B. In the preference manipulation, the
370 animal introduced themselves, then turned to one of the tables and expressed either that
371 they liked (“Oh wow, I really like that one”) or disliked (“Oh bleh, I really don’t like that
372 one”) the object before turning to the other side and expressing the respective other attitude.
373 Next the animal disappeared and, after a short pause, either the same or a different animal
374 returned and requested an object while facing straight ahead. This procedure was the strong
375 preference manipulation. In the medium version, the animal only expressed preference and
376 did so in a more subtle way (simply saying: “Oh, wow”).

377 In the novelty manipulation one of the tables was initially empty. The animal turned to
378 one of the sides and commented either on the presence (“Aha, look at that”) or the absence
379 (“Hm..., nothing there”) of an object before turning to the other side and commenting in a
380 complementary way. After shortly disappearing, the same animal repeated the sequence
381 above. When the animal left a second time, a new object appeared on the empty table. Next,
382 either the same or a different animal returned and requested an object. This corresponded to
383 the strong manipulation. For the medium manipulation, the animal turned to each table
384 only once before the new object appeared. In the weak manipulation, the animal only
385 commented on the present object once and never turned to the empty table. Participants

386 always received six trials, three with the same and three with the different speaker.

387 For experiment 3 and 4 we inserted the common ground manipulation before the
388 request in the setup of experiment 1 (Fig. 1C). For example, the animal turned to the table
389 with one object and express that they liked object A, then turned to the other table and
390 express that they did not like object B. Next, after quickly disappearing, the animal
391 reappeared, turned to the table with two objects and make a request. To make it clear,
392 which of the objects the speaker commented on while being turned to the table with the two
393 objects during the common ground manipulation, the object was temporarily enlarged.
394 Participants completed eight trials for one of the common ground manipulations with two
395 trials per condition (same/different speaker x congruent/incongruent).

396 Experiment 1 for children was modeled after Frank and Goodman (2014). Instead of
397 on tables, objects were presented as hanging in trees. After introducing themselves, the
398 animal turned to the tree with two objects and said: “This is a tree with a [non-word], how
399 neat, a tree with a [non-word]”). Next, the trees and the objects in them disappeared and
400 new trees replaced them. The two objects from the tree the animal turned to previously were
401 now spread across the two trees (one object per tree, position counterbalanced). While
402 facing straight, the animal first said “Here are some more trees” and then asked the child to
403 pick the tree with the object that corresponded to the novel word (“Which of these trees has
404 a [non-word]”). Children received six trials in a single test condition.

405 Experiment 2 for children was identical to the strong preference manipulation for
406 adults. Children received eight trials, four with the same and four with a different animal
407 returning.

408 In experiment 3 for children, we again inserted the preference manipulation into the
409 setup of experiment 1. After greeting the child, the animal turned to one of the trees,
410 pointed to an object (object was temporarily enlarged and moved closer to the animal) and
411 expressed liking or disliking. Then the animal turned to the other tree and expressed the
412 other attitude for the other kind of object. Next, the animal disappeared and either the

⁴¹³ same or a different animal returned. The rest of the trial was identical to the label and
⁴¹⁴ request phase of experiment 1. Children received eight trials, two per condition
⁴¹⁵ (same/different speaker x congruent/incongruent).

⁴¹⁶ **Analysis**

⁴¹⁷ All analysis were run in R (R Core Team, 2018). GLMMs were fit via the function
⁴¹⁸ `glmer` from the package `lme4` (Bates, Mächler, Bolker, & Walker, 2015). Probabilistic
⁴¹⁹ models and model comparisons were implemented in WebPPL (Goodman & Stuhlmüller,
⁴²⁰ 2014) using the r package `rwebppl` (Braginsky, Tessler, & Hawkins, 2019). The flat prior
⁴²¹ model had the same structure as the pragmatics model but used a uniform prior over objects.
⁴²² The prior only model predicted choice based on the prior probability for each object alone.
⁴²³ Note that the proportions that informed the prior distribution in the pragmatics and prior
⁴²⁴ only model were measured when participants chose between two objects in experiment 2. In
⁴²⁵ experiment 3, however, three objects were involved. For each object we used the proportion
⁴²⁶ measured in experiment 2 as the prior probability. This approach spread out the absolute
⁴²⁷ probability mass but conserved the relative relation between objects. Bayes Factors for
⁴²⁸ model comparisons are based on marginal likelihoods of each model given the data. The
⁴²⁹ corresponding model code can be found in the associated online repository.

References

- 431 Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017).
432 Bootstrapping language acquisition. *Cognition*, 164, 116–143.

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

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

437 Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the
438 communicative intentions behind gestures in a hiding game. *Developmental Science*,
439 8(6), 492–499.

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

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

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

446 Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first
447 language. In T. Moore (Ed.), *Cognitive development and acquisition of language* (pp.
448 65–110). New York: Academic Press.

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

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

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

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

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

- 457 speakers are informative. *Cognitive Psychology*, 75, 80–96.
- 458 Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1),
459 3–55.
- 460 Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic
461 inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- 462 Goodman, N. D., & Stuhlmüller, A. (2014). The design and implementation of probabilistic
463 programming languages. <http://dippl.org>.
- 464 Grice, H. P. (1991). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- 465 Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational
466 implicature*. Cambridge, MA: MIT press.
- 467 R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna,
468 Austria: R Foundation for Statistical Computing.
- 469 Saylor, M. M., Ganea, P. A., & Vázquez, M. D. (2011). What's mine is mine:
470 Twelve-month-olds use possessive pronouns to identify referents. *Developmental
471 Science*, 14(4), 859–864.
- 472 Saylor, M. M., Sabbagh, M. A., Fortuna, A., & Troseth, G. (2009). Preschoolers use
473 speakers' preferences to learn words. *Cognitive Development*, 24(2), 125–132.
- 474 Sperber, D., & Wilson, D. (2001). *Relevance: Communication and cognition* (2nd ed.).
475 Cambridge, MA: Blackwell Publishers.
- 476 Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool
477 children. *Language Learning and Development*, 11(2), 176–190.
- 478 Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT press.
- 479 Tomasello, M. (2009). *Constructing a language*. Cambridge, MA: Harvard University Press.