

1 How young children integrate information sources to infer the 2 meaning of words

3 Manuel Bohn^{1,*}, Michael Henry Tessler^{2,*}, Megan Merrick³, & Michael
4 C. Frank³

5 ¹ Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary
6 Anthropology

7 ² Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

8 ³ Department of Psychology, Stanford University

9 * These authors contributed equally to this work

10 Abstract

Before formal education begins, children typically acquire a vocabulary of
thousands of words. This learning process requires the use of many differ-
ent information sources in their social environment, including their current
knowledge state as well as the context in which they hear words used. How is
this information integrated? We specify a model according to which children
consider all available information sources and integrate them via Bayesian
inference. This model accurately predicted 2-to-5 year-old children’s word
learning across a range of experimental conditions. Model comparison sug-
gests that the integration process is stable across development and that the
main locus of development is an increased sensitivity to individual informa-
tion sources. We present a quantitative developmental theory of informa-
tion integration for language learning, embedded in a larger framework that
grounds language use and learning in social cognition.

Keywords: language acquisition, social cognition, pragmatics, Bayesian mod-
eling, common ground

Word count: X

12 Humans rely on language for communication, making language acquisition a key
13 learning objective of early childhood.¹ A central problem in language learning is referent
14 identification: To learn the conventional symbolic relation between a word and an object,
15 the child has to determine the intended referent of the word. Referents can only be identified
16 in an inferential manner by reasoning about the speaker’s intentions.²⁻⁵ That is, the child
17 has to infer what the speaker is communicating about based on information sources in the
18 social context of the utterance.

19 From early on in development, children use a number of different mechanisms to har-

ness such social-contextual information sources.^{2,4,6} For example, children expect speakers to use novel words for unknown objects,^{7–10} to talk about objects that are relevant,^{11,12} new in context,^{13,14} or related to the ongoing conversation.^{15–17} These different mechanisms, however, have largely been described and theorized about in isolation. The picture of the learning process that emerges is that of a “bag of tricks”: mechanisms that operate (and develop) independently from one another.⁶ As such, this view of the learning process does not capture the complexity of natural social interaction during which many sources of information are present.^{18,19} These sources need to be integrated in order to accurately infer a speaker’s intention.

When information integration is studied directly, the focus is mostly on how children interpret an ambiguous referring expression in light of social-contextual information.^{20–26} In one of these studies,²⁷ children faced a 2 x 2 display with a ball, a pen and two glasses in it. The speaker, sitting on the opposite side of the display, saw only 3 of the compartments: the ball, the pen, and one of the glasses. When the speaker asked for “the glass”, children had to integrate the semantics of the utterance with the perspective of the speaker to correctly infer which of the glasses - from their perspective - the speaker was referring to. This study, along with others, found that preschoolers use both information sources. What these studies do not do, however, is specify the process by which children integrate the different information sources. When discussing their results, authors refer to social-pragmatic theories of language use and learning,^{4,5,28–30} all of which assume that information is integrated as part of a social inference process, but none of which clearly defines the process. As a consequence, we have no explicit and quantitative theory of how different information sources (and word learning mechanisms) are integrated.

We present a theory of this integration process. Following social-pragmatic theories of language learning,^{4,5} our theory is based on the following premises: information sources serve different functional roles, all part of an integrated social inference process.^{28–31} Children use all available information to make inferences about the intentions behind a speaker’s utterance, which then leads them to correctly identify referents in the world and learn conventional word–object mappings. We formalize the computational steps that underlie this inference process in a cognitive model.^{32–34} In the remainder of this paper, we rigorously test this theory by asking how well it serves the two purposes of any psychological theory: prediction and explanation.^{35,36} First, we use the model to make quantitative predictions about children’s behavior in new situations - predictions we test against new data. Next, we test how well the model explains the integration process by comparing it to alternative models that make different assumptions about *whether* information is integrated, *how* it is integrated and how the integration process *develops*.

We focus on three information sources that play an integral part in theorizing about language use and learning: (1) expectations that speakers communicate in a cooperative and informative manner,^{7,11} (2) shared common ground about what is being talked about in conversation,^{30,37,38} and (3) semantic knowledge about previously learned word–object mappings.^{6,39} These information sources operate on different timescales: while (1) is a momentary expectation about a particular utterance, (2) grows over the course of a conversation, and (3) is learned across development. This interplay of timescales has been hypothesized to be an important component of word meaning inference.^{40,41}

Our *rational integration model* arbitrates between information sources via Bayesian inference (see Fig. 1f for model formula). A listener (L_1) reasons about the referent of a speaker's (S_1) utterance. This reasoning is contextualized by the prior probability of each referent ρ – which we take to be a function of the common ground shared between the listener and the speaker. To decide between referents, the listener (L_1) reasons about what a rational speaker (S_1) with informativeness α would say given an intended referent. This speaker is assumed to compute the informativity for each available utterance and then choose the most informative one. The informativity of each utterance is given by imagining which referent a listener, who interprets words according to their literal semantics (what we call a literal listener, L_0), would infer upon hearing the utterance. Naturally, this reasoning depends on what kind of semantic knowledge (for object j) θ_j the speaker ascribes to the (literal) listener. Taken together, this model provides a quantitative theory of information integration during language learning. Furthermore, the model presents an explicit and substantive theory of development. It assumes that, while children's sensitivity to the individual information sources increases with age, the way integration proceeds remains constant.^{2,42} In the model, this is accomplished by making the parameters capturing sensitivity to speaker informativeness (α_i , Fig. 1d), the common ground (ρ_i , Fig. 1c), and object specific semantic knowledge ($\theta_{i,j}$, Fig. 1e) a function of the child's age.

To test the predictive and explanatory power of our model, we developed a world-learning paradigm in which we jointly manipulated the three information sources (Fig. 1). Children interacted on a tablet computer with a series of storybook speakers.⁴³ The speakers used novel words (e.g., “wug”) in the context of two potential referents. This situation is depicted in Fig. 1a iv, in which a speaker (here, a frog) appears with a known object (a duck, left) and an unfamiliar object (the diamond-shaped object, right). To identify the referent, the child could rely upon the three information sources described above. First, they may infer that a cooperative speaker would have used the word “duck” to refer to the known object (the duck); the fact that the speaker did not say duck then suggests that the speaker is most likely referring to a different object (the unfamiliar object). This basic inference is oftentimes referred to as a mutual exclusivity inference.^{8,10} More generally speaking, the listener can infer the referent of the novel word by assuming that the speaker is a rational agent who chooses the most informative words to refer to objects.^{7,11} Second, the child may rely upon what has already been established in the common ground with the speaker. Listeners expect speakers to communicate about things that are new to the common ground.^{13,14} Thus, the inference about the novel word referring to the unfamiliar object also depends on which object is new in context (Fig. 1a and b i-iii). Finally, the child may rely upon their previously acquired semantic knowledge, that is, how sure they are that the known object is called “duck”. If the known object is something less familiar, such as a chess piece (e.g., a pawn), a 3-year-old child may draw a weaker inference, if they draw any inference at all.⁴⁴ Taken together, the child has the opportunity to integrate their assumptions about (1) informative communication, (2) their understanding of the common ground, and (3) their existing semantic knowledge. In some versions of the experiment, some of these information sources were aligned (Fig. 1a) but in some they were in conflict (Fig. 1b).

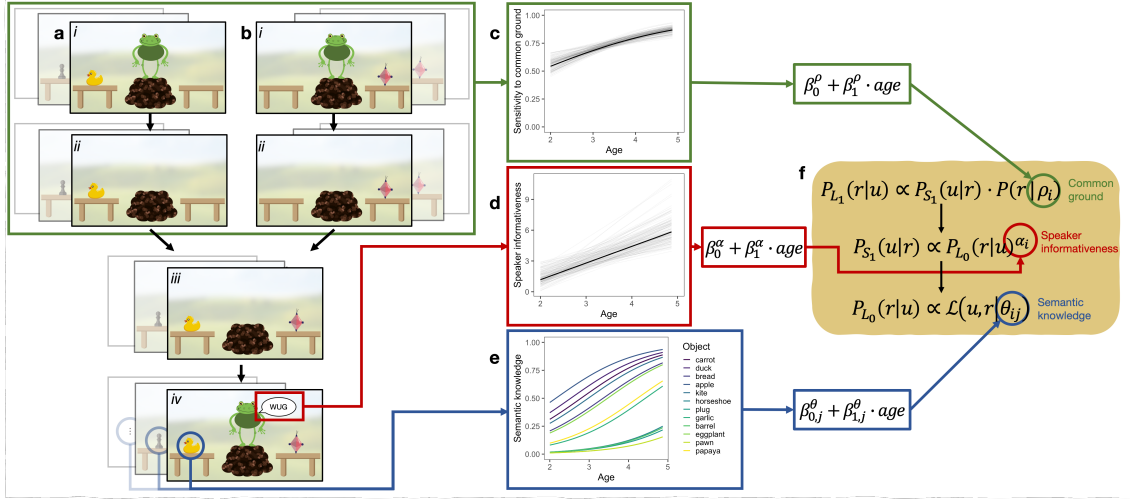


Figure 1. Experimental task and model. (a and b) Screenshots from the experimental task. (i) The speaker encounters one object and then leaves the scene (ii). While the speaker is away, a second object appears (iii), when returning, the speaker uses a novel word to request an object (iv). Sections (i) to (iii) establish common ground between the speaker and the listener, in that one object is new in context (green). The request in (iv) licenses an inference based on expectations about how informative speakers are (red). Listeners' semantic knowledge enters the task because the identity of the known object on one of the tables is varied from well-known objects like a duck to relatively unfamiliar objects like a chess pawn (total of 12 objects - blue). (a) shows the condition in which common ground information is congruent (i.e. point to the same object) with speaker informativeness and (b) shows the incongruent condition. The congruent and incongruent conditions are each paired with the 12 known objects, resulting in 24 unique conditions. (c to e) show developmental trajectories for each information source, estimated from separate experiments (see supplement). (f) gives the model equation for the rational integration model and links information sources to model parameters.

Results

Predicting information integration across development

We tested the model in its ability to predict children's actual behavior - at an age (between 2 and 5 years) at which they learn language at an exponential rate.⁶ We estimated children's (N=148) developing sensitivity to individual information sources in separate experiments (Experiment 1 and 2) (see Fig. 1c-e). We then generated parameter-free, quantitative, *a priori* model predictions (developmental trajectories) representing the model's expectations about how children should behave in a new situation in which all three information sources had to be integrated. We generated model predictions for 24 experimental conditions: 12 objects of different familiarities (requiring different levels of semantic knowledge), with novelty either conflicting or coinciding; Fig. 1. We compared these predictions to newly collected data from N = 220 children from the same age range (Experiment 3). The results showed a very close alignment between model predictions and the data across

the entire age range. That is, the average developmental trajectories predicted by the model resembled the trajectories found in the data (Fig. S6). For a more quantitative analysis, we binned predictions and data by year and correlated them. We found a high correlation, with the model explaining 79% of the variance in the data (Fig. 2a). These results seem to support the assumptions in the model that children integrate all available information sources. However, it is still possible that simpler models might make equally good – or even better – predictions.

Thus, we formalized the alternative view that children selectively ignore information sources in the form of three lesioned models (Fig. 2b). These models assume that children follow the heuristic “ignore X” (with X being one of the information sources) when multiple information sources are presented together.

The *no word knowledge model* uses the same model architecture as the rational integration model. It takes in expectations about speaker informativeness and common ground, but omits semantic knowledge that is specific to the familiar objects (i.e., uses only general semantic knowledge). That is, the model assumes a listener whose inference does not vary depending on the particular familiar object but only depends on the age specific average semantic knowledge. The *no common ground model* takes in object specific semantic knowledge and speaker informativeness but ignores common ground information. Instead of assuming that one object has a higher prior probability to be the referent because it is new in context, the speaker thinks that both objects are equally likely to be the referent. As a consequence, the listener does not differentiate between situations in which common ground is aligned or in conflict with the other information sources. Finally, according to the *no speaker informativeness model*, the listener does not assume that the speaker is communicating in an informative way and hence ignores the information provided by the utterance. As a consequence, the inference is solely based on common ground expectations.

We found little support that children make inferences following one of the heuristic models (Fig. 2b). When using Bayesian model comparison via marginal likelihood of the data,⁴⁵ we find that the data was several orders of magnitude more likely under the rational integration model compared to any of the lesioned models (Fig. 2). Taken together, we may conclude that children considered all available information sources.

Explaining the process of information integration

The fact that children integrated all the available information sources may be surprising, but does not rule out that the integration process itself might be different from the one described in the rational integration model. Remember that this model assumes that all information sources enter into an integrated inference process. However, a more modular view seems equally plausible: children might compute separate inferences based on a subset of the available information and then integrate them by weighting them according to some ratio. This view would allow for the possibility that some information sources are considered to be more important than others. In other words, children might be biased towards some information sources. We formalized this alternative view as a *biased integration model*. This model assumes that semantic knowledge and expectations about speaker informativeness enter into one inference (mutual exclusivity inference^{7,8,44}) while

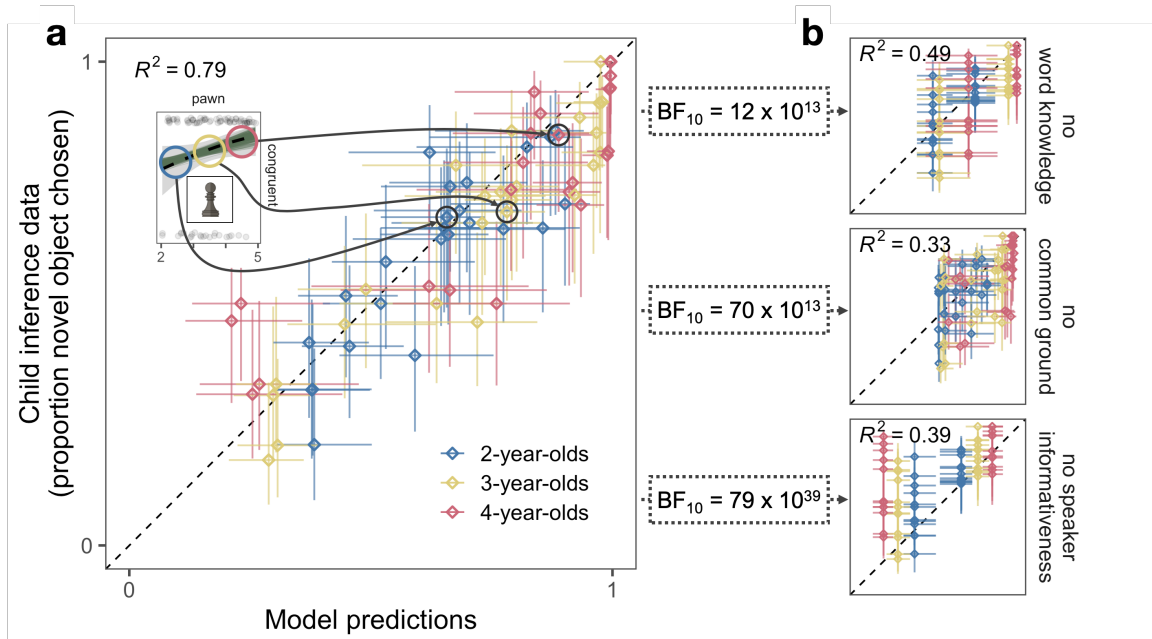


Figure 2. Predicting information integration. Correlation between model predictions and child inference data for all 24 conditions and for each age group (binned by year) for the rational integration model (a) and the three lesioned models (b). Horizontal and vertical error bars show 95% HDI. Inset shows an example of model predictions as developmental trajectories (see Fig. 3). The no word knowledge model includes developmental change for overall vocabulary development but ignores the identity of the known object. The no common ground model assumes that children ignore information from common ground and the no speaker informativeness model assumes that children ignore inferences that are licensed by expectations about informative communication. BF_{10} gives the Bayes Factor in favor of the integration model based on the marginal likelihood of the data under each model.

162 common ground information enters into a second one. The outcome of both processes are
 163 then weighted according to the parameter ϕ . Like the rational integration model, this model
 164 takes in all available information sources in an age sensitive way and assumes that they are
 165 integrated - the integration process itself, however, is construed in a different way.

166 The parameter ϕ in the biased integration model is unknown ahead of time and has to
 167 be estimated based on the experimental data. That is, in contrast to the rational integration
 168 model, the biased integration model does not allow to make *a priori* predictions about new
 169 data in the way we describe above. For the model comparison, we therefore constrained
 170 the parameters in the rational integration model by the experimental data as well. As a
 171 consequence, both models use all the information available to constrain their parameters
 172 (see Fig. S4).

173 The biased integration model makes reasonable predictions that explain a large pro-
 174 portion of the variance in the data (Fig. 3b). The parameter ϕ - indicating the bias to one of
 175 the inferences - was estimated to favor the mutual exclusivity inference (mode = 0.65; 95%

highest density interval (HDI): 0.60 - 0.71, see Fig. 3d). However, the rational integration model presented a much better fit to the data, both in terms of correlation and marginal likelihood of the data (Fig. 3). Thus, a fully integrated - as opposed to a segmented and biased - integration process explained the data better.

The rational integration model assumes that the integration process itself does not change with age.² That is, while children’s sensitivity to each information source develops, the way they relate to one another remains the same. An alternative perspective would be to assume that the integration process itself changes with age. The biased integration model provides a way to think about developmental change in the integration process: Children are biased towards some information sources, but the bias itself changes with age. We formalize such an alternative view as a *developmental bias model* which is structurally identical to the biased integration model but in which the parameter ϕ changes with age. The model assumes that the importance of the different information sources changes with age.

The developmental bias model explains a substantial portion of the variance in the data (Fig. 3c). The estimated developmental trajectory for the bias parameter ϕ suggests that younger children put a higher emphasis on common ground information, while older children rely more on the mutual exclusivity inference (Fig. 3d). The relative importance of the two inferences seems to switch at around age 3. Yet again, when we directly compare the competitor models, we find that the data is several orders of magnitude more likely under the rational integration model (Fig. 3). Interestingly, when we compared the rational integration model with its parameters not constrained by the experimental data (i.e. the “pure” prediction model described above), it also outperformed both bias models (see Fig. S10 and Tab. S6). Taken together, these results show that the integration process is best described as fully integrated and stable over time.

Discussion

The environment in which children learn language is complex. Children have to integrate different information sources, some of which relate to expectations in the moment, others to the dynamics of the unfolding interactions and yet others to their previously acquired knowledge. Our findings show that young children can flexibly integrate multiple information sources during language learning - even from relatively early in development. To answer the question of how they do so, we presented a formal cognitive model that assumes that information sources are rationally integrated via Bayesian inference.

Our model is based on classic social- pragmatic views on language use and comprehension.^{5,28–30} As a consequence, instead of assuming that different information sources feed into separate word-learning mechanisms, we assume that they all play a functional role in an integrated social inference process. Our model goes beyond previous theoretical and empirical work by specifying the computations that may underlie this inference process. Furthermore, we present a substantive theory about how this integration process develops: We assume that children become increasingly sensitive to different information sources but that the way these information sources are integrated, remains the same. We used this model to predict and explain children’s information integration in a new word

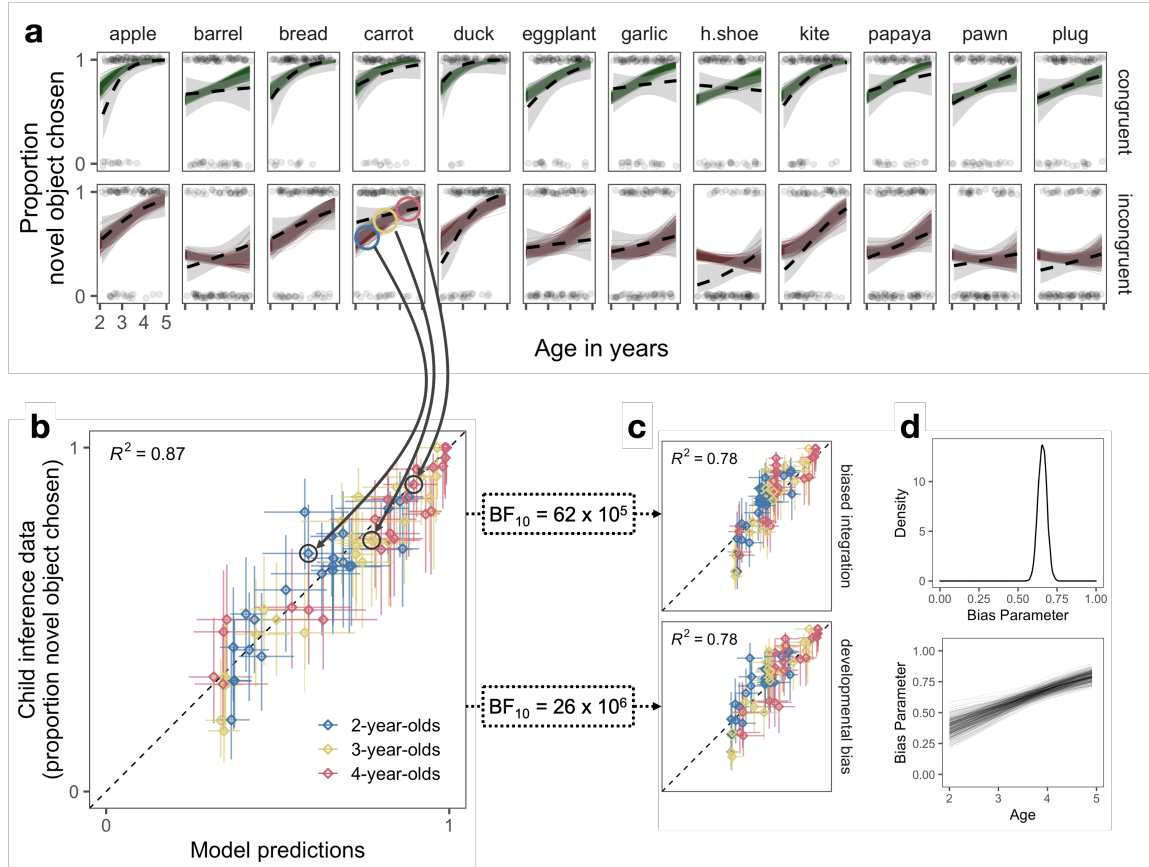


Figure 3. Explaining information integration across development. (a) Model predictions from the rational integration model (colored lines) next to the behavioral data (dotted black lines with 95% CI in gray) for all 24 experimental conditions. Top row (blue) shows conditions in which common ground information is congruent with speaker informativeness, bottom row (red) shows conditions in which they are incongruent. Known objects are ordered based on their rated age of acquisition (left to right). Light dots represent individual data points. (b) Correlations between model predictions binned by age and condition for the integration model and (c) the two mixture models. Vertical and horizontal error bars show 95% HDIs. BF_{10} gives the Bayes Factor in favor of the rational integration model based on the marginal likelihood of the data under each model. (d) Posterior distribution of the bias parameter in the biased integration model and developmental trajectories for the bias parameter in the developmental bias model based on 300 random draws from the posterior distribution of the intercept and slope.

learning paradigm in which they had to integrate (1) their assumptions about informative communication, (2) their understanding of the common ground, and (3) their existing semantic knowledge.

We found that this rational integration model made accurate quantitative predictions across a range of experimental conditions - in some of which information sources were aligned, in some of which they were in conflict. Predictions from the model better explained the data compared to lesioned models which assumed that children ignore one of the information sources, suggesting that children used all available information.

To test the explanatory power of the model - how well it explains the process by which information is integrated - we formalized an alternative, modular, view. According to the biased integration model, children use all available information sources but compute separate inferences based on a subset of them. Integration happens by weighting the outcomes of these separate inferences by some ratio. By allowing the ratio to be in favor of one of the inferences, this model captures the idea that children might be biased to some information sources over others. Even though such a model made reasonable predictions, we found that the rational integration model explained children's behavior much more accurately. Finally, we tested an alternative view on the development of the integration process. According to the developmental bias model, the importance of the different information sources changed with age. Again, the rational integration model provided a much better fit to the data, suggesting that the integration process remains stable over time.

The rational integration model is derived from a more general framework for pragmatic inference, which has been used to explain a wide variety of phenomena in adults' language use and comprehension.^{32,33,46-49} Thus, it can be generalized in a natural way to capture word learning in contexts that offer more, fewer, or different types of information. For example, in a recent study with adults and children, Bohn and colleagues⁴² manipulated expectations about informativeness and common ground in a different way. Using a structurally similar model, they also found a close alignment between model predictions and the data. The flexibility of this modeling framework stems from its conceptualization of human communication as a form of rational social action. As such, it connects to computational and empirical work that tries to explain social reasoning by assuming that humans expect each other to behave in a way that maximizes the benefits and minimizes the cost associated with actions.^{23,50,51}

There are multiple ways in which the model and the experiments presented here could be extended. First, it would be important to replicate our findings in a more diverse sample, that is, with children from a different cultural background who learn a different language. In such a study, we would not expect that our results replicate in a strict sense; that is, we would not expect to see the same developmental trajectories in all cultures and/or languages. Substantial variation is much more likely. Our computational model provides a way to think about how to reconcile cross-cultural variation with a universal cognitive architecture: We expect differences in how sensitive children are to the individual information sources at different ages, but similarities in how information is integrated.² In computational terms, we assume a universal architecture that specifies the relation between a set of varying parameters.

Second, it would be useful to flesh out the cognitive processes that underlie reasoning about common ground. In its current form, our model does not specify how the listener arrives at the expectation that some objects are more likely referents because they are new in common ground. We assume that common ground changes speaker’s expectations about what are likely referents⁵² and measured these expectations to inform the parameter values in our model. Computationally, the model does not differentiate between common ground information and other reasons that might make an object contextually more salient. A good starting point to address this issue could be work that formalizes reasoning about common ground as trading-off between (differing) perspectives.^{53,54}

Finally, our model is a model of referent identification in the moment of the utterance. Word learning, however, is a cumulative process which happens across multiple time points and contexts. Thus, it would be fruitful to connect our model with frameworks that focus on the temporal and cross-situational aspects of word learning.^{40,41,55}

Taken together, this work helps us understand how children navigate the complexity of their learning environment. Methodologically, it illustrates how formal, computational models can be used to test theories - bot in terms of their predictive and explanatory power. From a theoretical perspective, it adds to broader frameworks that see the onto- and phylogenetic emergence of language as deeply rooted in social cognition.

Methods

A more detailed description of the experiments and the models can be found in the supplementary material. The experimental procedure, sample size and analysis for each experiment were pre-registered (<https://osf.io/7rg9j/registrations>). Experimental procedures, model and analysis scripts can be found in an online repository (<https://github.com/manuelbohn/spin>). Experiment 1 and 2 were designed to estimate children’s developing sensitivity to each information source. Their results inform the parameter values in the model (see Fig. 1 c-f). Experiment 3 was designed to test how children integrate the different information sources.

Participants

Across the three experiments, a total of 368 children participated. Experiment 1 involved 90 children, including 30 2-year-olds (range = 2.03 - 3.00, 15 girls), 30 3-year-olds (range = 3.03 - 3.97, 22 girls) and 30 4-year-olds (range = 4.03 - 4.90, 16 girls). Data from 10 additional children was not included because they were either exposed to less than 75% of English at home (5), did not finish at least half of the test trials (2), the technical equipment failed (2) or their parents reported an autism spectrum disorder (1).

In Experiment 2, we tested 58 children from the same general population as in Experiment 1, including 18 2-year-olds (range = 2.02 - 2.93, 7 girls), 19 3-year-olds (range = 3.01 - 3.90, 14 girls) and 21 4-year-olds (range = 4.07 - 4.93, 14 girls). Data from 5 additional children was not included because they were either exposed to less than 75% of English at home (3) or the technical equipment failed (2).

Finally, Experiment 3 involved 220 children, including 76 2-year-olds (range = 2.04 - 2.99, 7 girls), 72 3-year-olds (range = 3.00 - 3.98, 14 girls) and 72 4-year-olds (range = 4.00 - 4.94, 14 girls). Data from 20 additional children was not included because they were either exposed to less than 75% of English at home (15), did not finish at least half of the test trials (3) or the technical equipment failed (2).

All participants were recruited from the floor of a Children’s museum in San José, California, USA. This population is characterized by diverse ethnic background (predominantly White, Asian, or mixed ethnicity) and high levels of parental education and socioeconomic status. Parents consented to their children’s participation and provided demographic information. All experiments were approved by the Stanford Institutional Review Board (protocol no. 19960).

Materials

All experiments were presented as an interactive picture book on a tablet computer.⁴³ Fig. 1a and b show screenshots from the actual experiments. The general setup involved an animal standing on a little hill between two tables. For each animal character, we recorded a set of utterances (one native English speaker per animal) that were used to talk to the child and make requests. Each experiment started with two training trials in which the speaker requested known objects (car and ball).

Procedure

Experiment 1 tested the so-called mutual exclusivity inference.^{8,44} On one table, there was a familiar object, on the other table, there was an unfamiliar object (drawn for the purpose of the study) (Fig. 1a/b iv and Fig. S1a). The speaker requested an object by saying “Oh cool, there is a [non-word] on the table, how neat, can you give me the [non-word]?”. Children responded by touching one of the objects. The location of the unfamiliar object (left or right table) and the animal character were counterbalanced. We coded as correct choice if children chose the unfamiliar object as the referent of the novel word. Each child completed 12 trials, each with a different familiar and a different unfamiliar object. Familiar objects were selected to vary along the dimension of how likely children were to know the word for it. This included objects that most 2-year-olds can name (e.g. a duck) as well as objects that only very few 5-year-olds can name (e.g. a pawn). The selection was based on age of acquisition ratings from Kuperman and colleagues.⁵⁶ While these ratings do not capture the absolute age when children acquire these words, they capture the relative order in which words are learned. Fig. S2A in the supplementary material shows the words and objects used in the experiment.

Experiment 2 tested children’s sensitivity to common ground that is build up over the course of a conversation. In particular, we tested whether children keep track of which object is new to a speaker and which they have encountered previously.^{13,14} The general setup was the same as in Experiment 1 (Fig. S1b). The speaker was positioned between the tables. There was an unfamiliar object (drawn for the purpose of the study) on one of the tables while the other table was empty. Next, the speaker turned to one of the tables

and either commented on the presence (“Aha, look at that.”) or the absence (“Hm, nothing there”) of an object. Then the speaker disappeared. While the speaker was away, a second unfamiliar object appeared on the previously empty table. Then the speaker returned and requested an object in the same way as in Experiment 1. The positioning of the unfamiliar object in the beginning of the experiment, the speaker as well as the location the speaker turned to first was counterbalanced. Children received five trials, each with a different pair of unfamiliar objects. We coded as correct choice if children chose the object that was new to the speaker as the referent of the novel word.

Experiment 3 combined the procedures from Experiment 1 and 2. It followed the same procedure as Experiment 2 but involved the same objects as Experiment 1 (Fig. 1 i-iv and Fig. S1c). In the beginning, one table was empty while there was an object (unfamiliar or familiar) on the other one. After commenting on the presence or absence of an object on each table, the speaker disappeared and a second object appeared (familiar or unfamiliar). Next, the speaker re-appeared and made the usual request. In the congruent condition, the familiar object was present in the beginning and the unfamiliar object appeared while the speaker was away (Fig. 1a and Fig. S1c - left). In this case, both the mutual exclusivity and the common ground inference pointed to the novel object as the referent. In the incongruent condition, the unfamiliar object was present in the beginning and the familiar object appeared later. In this case, the two inferences pointed to different objects (Fig. 1b and Fig. S1c - right). This resulted in a total of 2 (congruent vs incongruent) \times 12 (familiar objects) = 24 different conditions. Participants received up to 12 test trials, six in each alignment condition, each with a different familiar and unfamiliar object. Familiar objects were the same as in Experiment 1. The positioning of the objects on the tables, the speaker and the location the speaker first turned to were counterbalanced. Participants could stop the experiment after six trials (three per alignment condition). If a participant stopped after half of the trials, we tested an additional participant to reach the preregistered number of data points per age group (2-, 3- and 4-year-olds).

Data analysis

To analyze how the manipulations in each experiment affected children’s behavior, we used generalized linear mixed models. Since the focus of the paper is on how information sources were integrated, we discuss these models in the supplementary material and focus here on the cognitive models instead. A detailed, mathematical description of the different cognitive models along with details about estimation procedures and priors can be found in the supplementary material. All Bayesian cognitive models were implemented in the probabilistic programming language WebPPL.⁵⁷ The corresponding model code can be found in the associated online repository.

As a first step, we used the data from Experiment 1 and 2 to estimate children’s developing sensitivity to each information source. To estimate the parameters for semantic knowledge (θ) and speaker informativeness (α), we adapted the rational integration model to a situation in which both objects (novel and familiar) have equal prior probability (i.e., no common ground information). We used the data from Experiment 1 to then infer the parameters. That is, we inferred the intercepts and slopes for speaker informativeness (linear

regression) and semantic knowledge (logistic regression) that generated model predictions to match the responses generated in Experiment 1. To estimate the parameters representing sensitivity to common ground (ρ), we used a simple logistic regression to infer which combination of intercept and slope would generate predictions that corresponded to the average proportion of correct responses measured in Experiment 2. For the explanation models, these parameters were additionally constrained by the data from Experiment 3. A more detailed description of how these parameters were estimated (including a graphical model) can be found in the supplementary material.

To generate model predictions, we combined the parameters according to the respective model formula. As mentioned above, common ground information could either be aligned or in conflict with the other information sources. In the congruent condition, the unfamiliar object was also new in context and thus had the prior probability ρ . In the incongruent condition, the novel object was the “old” object and thus had the prior probability of $1 - \rho$.

The rational integration model is defined as $P_{L_1}^{int}(r | u; \{\rho_i, \alpha_i \theta_{ij}\}) \propto P_{S_1}(u | r; \{\alpha_i, \theta_{ij}\}) \cdot P(r | \rho_i)$. The three lesioned models that were used to compare how well the model predicts new data are reduced versions of this model. The no word knowledge model uses the same model architecture: $P_{L_1}^{no-wk}(r | u; \{\rho_i, \alpha_i \theta_{ij}\}) \propto P_{S_1}(u | r; \{\alpha_i, \theta_{ij}\}) \cdot P(r | \rho_i)$ and the only difference lies in the parameter θ . The object specific parameters for semantic knowledge are fitted via a hierarchical regression (mixed effects) model. That is, there is an overall developmental trajectory for semantic knowledge (main effect - θ_i) and then there is object-specific variation around this trajectory (random effects - θ_{ij}). Thus, the no word knowledge model takes in the overall trajectory for semantic knowledge (θ_i) but ignores object-specific variation. The no common ground model ignores common ground information (represented by ρ) and is thus defined as $P_{L_1}^{no-cg}(r | u; \{\alpha_i \theta_{ij}\}) \propto P_{S_1}(u | r; \{\alpha_i, \theta_{ij}\})$. For the no speaker informativeness model, the parameter $\alpha = 0$. As a consequence, the likelihood term in the model is 1 and the model therefore reduces to $P_{L_1}^{no-si}(r | u; \{\rho_i\}) \propto P(r | \rho_i)$.

As noted above, the explanation models used parameters that were additionally constrained by the data from Experiment 3. This had no effect on the way these parameters were combined in the rational integration model. The biased integration model is defined as $P_{L_1}^{biased}(r | u; \{\phi, \rho_i, \alpha_i, \theta_{ij}\}) = \phi \cdot P_{ME}(r | u; \{\alpha_i, \theta_{ij}\}) + (1 - \phi) \cdot P(r | \rho_i)$ with P_{ME} representing a mutual exclusivity inference which takes in speaker informativeness and object specific semantic knowledge. This inference is then weighted by the parameter ϕ and added to the respective prior probability, which is weighted by $1 - \phi_i$. Thus, ϕ represents the bias in favor of the mutual exclusivity inference. In the developmental bias model the parameter ϕ is made to change with age (ϕ_i) and the model is thus defined as $P_{L_1}^{dev-bias}(r | u; \{\phi_i, \rho_i, \alpha_i, \theta_{ij}\}) = \phi_i \cdot P_{ME}(r | u; \{\alpha_i, \theta_{ij}\}) + (1 - \phi_i)$.

We compared models in two ways. First, we used Pearson correlations between model predictions and the data. For this analysis, we binned the model predictions and the data by age in years and by the type of familiar object (see Fig. 2 and 3 as well as S7 and S10). Second, we compared models based on the marginal likelihood of the data under each model – the likelihood of the data averaging over (“marginalizing over”) the prior

424 distribution on parameters; the pairwise ratio of marginal likelihoods for two models is
425 known as the Bayes Factor. It is interpreted as how many times more likely the data is
426 under one model compared to the other. Bayes Factors quantify the quality of predictions
427 of a model, averaging over the possible values of the parameters of the models (weighted by
428 the prior probabilities of those parameter values); by averaging over the prior distribution
429 on parameters, Bayes Factor implicitly take into account model complexity because models
430 with more parameters will tend to have a broader prior distribution over parameters, which
431 in effect, can water down the potential gains in predictive accuracy that a model with more
432 parameters can achieve.⁴⁵ For this analysis, we treated age continuously.

References

1. Tomasello, M. *A natural history of human thinking*. (Harvard University Press, 2018).
2. Bohn, M. & Frank, M. C. The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology* **1**, 223–249 (2019).
3. Bruner, J. *Child's talk: Learning to use language*. (New York: Norton, 1983).
4. Clark, E. V. *First language acquisition*. (Cambridge: Cambridge University Press, 2009).
5. Tomasello, M. *Constructing a language*. (Cambridge, MA: Harvard University Press, 2009).
6. Bloom, P. *How children learn the meanings of words*. (MIT press, 2002).
7. Clark, E. V. The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition* (1987).
8. Markman, E. M. & Wachtel, G. F. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology* **20**, 121–157 (1988).
9. Carey, S. & Bartlett, E. Acquiring a single new word. *Proceedings of the Stanford Child Language Conference* **15**, 17–29 (1978).
10. Halberda, J. The development of a word-learning strategy. *Cognition* **87**, B23–B34 (2003).
11. Frank, M. C. & Goodman, N. D. Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology* **75**, 80–96 (2014).
12. Schulze, C., Grassmann, S. & Tomasello, M. 3-year-old children make relevance inferences in indirect verbal communication. *Child Development* **84**, 2079–2093 (2013).
13. Akhtar, N., Carpenter, M. & Tomasello, M. The role of discourse novelty in early word learning. *Child Development* **67**, 635–645 (1996).
14. Diesendruck, G., Markson, L., Akhtar, N. & Reudor, A. Two-year-olds' sensitivity to speakers' intent: An alternative account of samuelson and smith. *Developmental Science* **7**, 33–41 (2004).
15. Bohn, M., Le, K., Peloquin, B., Koymen, B. & Frank, M. C. Children's interpretation of ambiguous pronouns based on prior discourse. *PsyArXiv* (2020). doi:10.31234/osf.io/gkhez
16. Sullivan, J. & Barner, D. Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental Science* **19**, 63–75 (2016).
17. Horowitz, A. C. & Frank, M. C. Young children's developing sensitivity to discourse continuity as a cue for inferring reference. *Journal of experimental child psychology* **129**, 84–97 (2015).
18. Bergelson, E. & Aslin, R. N. Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences* **114**, 12916–12921 (2017).

- 469 19. Frank, M. C., Tenenbaum, J. B. & Fernald, A. Social and discourse contributions to
470 the determination of reference in cross-situational word learning. *Language Learning*
471 *and Development* **9**, 1–24 (2013).
- 472 20. Ganea, P. A. & Saylor, M. M. Infants’ use of shared linguistic information to clarify
473 ambiguous requests. *Child Development* **78**, 493–502 (2007).
- 474 21. Graham, S. A., San Juan, V. & Khu, M. Words are not enough: How preschoolers’ inte-
475 gration of perspective and emotion informs their referential understanding. *Journal*
476 *of Child Language* **44**, 500–526 (2017).
- 477 22. Grosse, G., Moll, H. & Tomasello, M. 21-month-olds understand the cooperative logic
478 of requests. *Journal of Pragmatics* **42**, 3377–3383 (2010).
- 479 23. Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B. & Schulz, L. E. Social prag-
480 matics: Preschoolers rely on commonsense psychology to resolve referential under-
481 specification. *Child Development* **91**, 1135–1149 (2020).
- 482 24. Khu, M., Chambers, C. G. & Graham, S. A. Preschoolers flexibly shift between speakers’
483 perspectives during real-time language comprehension. *Child development* **91**, e619–
484 e634 (2020).
- 485 25. Matthews, D., Lieven, E., Theakston, A. & Tomasello, M. The effect of perceptual
486 availability and prior discourse on young children’s use of referring expressions.
487 *Applied Psycholinguistics* **27**, 403–422 (2006).
- 488 26. Nilsen, E. S., Graham, S. A. & Pettigrew, T. Preschoolers’ word mapping: The interplay
489 between labelling context and specificity of speaker information. *Journal of child*
490 *language* **36**, 673–684 (2009).
- 491 27. Nadig, A. S. & Sedivy, J. C. Evidence of perspective-taking constraints in children’s
492 on-line reference resolution. *Psychological Science* **13**, 329–336 (2002).
- 493 28. Grice, H. P. *Studies in the way of words*. (Cambridge, MA: Harvard University Press,
494 1991).
- 495 29. Sperber, D. & Wilson, D. *Relevance: Communication and cognition*. (Cambridge, MA:
496 Blackwell Publishers, 2001).
- 497 30. Clark, H. H. *Using language*. (Cambridge: Cambridge University Press, 1996).
- 498 31. Shafto, P., Goodman, N. D. & Frank, M. C. Learning from others: The consequences of
499 psychological reasoning for human learning. *Perspectives on Psychological Science*
500 **7**, 341–351 (2012).
- 501 32. Frank, M. C. & Goodman, N. D. Predicting pragmatic reasoning in language games.
502 *Science* **336**, 998–998 (2012).
- 503 33. Goodman, N. D. & Frank, M. C. Pragmatic language interpretation as probabilistic
504 inference. *Trends in Cognitive Sciences* **20**, 818–829 (2016).
- 505 34. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psycho-*
506 *nomic bulletin & review* **26**, 1596–1618 (2019).

- 507 35. Shmueli, G. & others. To explain or to predict? *Statistical Science* **25**, 289–310 (2010).
- 508 36. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons
509 from machine learning. *Perspectives on Psychological Science* **12**, 1100–1122 (2017).
- 510 37. Bohn, M. & Koymen, B. Common ground and development. *Child Development Per-*
511 *spectives* **12**, 104–108 (2018).
- 512 38. Clark, E. V. Common ground. in *The handbook of language emergence* (eds. MacWhin-
513 ney, B. & O’Grady, W.) **87**, 328–353 (John Wiley & Sons, 2015).
- 514 39. Fenson, L. *et al.* Variability in early communicative development. *Monographs of the*
515 *Society for Research in Child Development* i–185 (1994).
- 516 40. Frank, M. C., Goodman, N. D. & Tenenbaum, J. B. Using speakers’ referential intentions
517 to model early cross-situational word learning. *Psychological Science* **20**, 578–585
518 (2009).
- 519 41. McMurray, B., Horst, J. S. & Samuelson, L. K. Word learning emerges from the
520 interaction of online referent selection and slow associative learning. *Psychological*
521 *Review* **119**, 831 (2012).
- 522 42. Bohn, M., Tessler, M. H., Merrick, M. & Frank, M. C. Predicting pragmatic cue
523 integration in adults’ and children’s inferences about novel word meanings. (2019).
524 doi:10.31234/osf.io/xma4f
- 525 43. Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L. & Yurovsky, D. Using
526 tablets to collect data from young children. *Journal of Cognition and Development*
527 **17**, 1–17 (2016).
- 528 44. Lewis, M. L., Cristiano, V., Lake, B. M., Kwan, T. & Frank, M. C. The role of
529 developmental change and linguistic experience in the mutual exclusivity effect.
530 *Cognition* **198**, 104191 (2020).
- 531 45. Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: A practical course*.
532 (Cambridge University Press, 2014).
- 533 46. Monroe, W., Hawkins, R. X., Goodman, N. D. & Potts, C. Colors in context: A
534 pragmatic neural model for grounded language understanding. *Transactions of the*
535 *Association for Computational Linguistics* **5**, 325–338 (2017).
- 536 47. Wang, S., Liang, P. & Manning, C. D. Learning language games through interaction.
537 *Proceedings of the 54th Annual Meeting of the Association for Computational Lin-*
538 *guistics* 2368–2378 (2016).
- 539 48. Tessler, M. H. & Goodman, N. D. The language of generalization. *Psychological Review*
540 **126**, 395–436 (2019).
- 541 49. Yoon, E. J., Frank, M. C., Tessler, M. H. & Goodman, N. D. Polite speech emerges
542 from competing social goals. *psyarxiv.com/67ne8* (2018).
- 543 50. Bridgers, S., Jara-Ettinger, J. & Gweon, H. Young children consider the expected utility
544 of others’ learning to decide what to teach. *Nature Human Behaviour* **4**, 144–152
545 (2020).

- 546 51. Jara-Ettinger, J., Gweon, H., Schulz, L. E. & Tenenbaum, J. B. The naive utility
547 calculus: Computational principles underlying commonsense psychology. *Trends in*
548 *Cognitive Sciences* **20**, 589–604 (2016).
- 549 52. Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. The effects of common ground
550 and perspective on domains of referential interpretation. *Journal of Memory and*
551 *Language* **49**, 43–61 (2003).
- 552 53. Heller, D., Parisien, C. & Stevenson, S. Perspective-taking behavior as the probabilistic
553 weighing of multiple domains. *Cognition* **149**, 104–120 (2016).
- 554 54. Mozuraitis, M., Stevenson, S. & Heller, D. Modeling reference production as the proba-
555 bilistic combination of multiple perspectives. *Cognitive science* **42**, 974–1008 (2018).
- 556 55. Yurovsky, D. & Frank, M. C. An integrative account of constraints on cross-situational
557 learning. *Cognition* **145**, 53–62 (2015).
- 558 56. Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings
559 for 30,000 english words. *Behavior Research Methods* **44**, 978–990 (2012).
- 560 57. Goodman, N. D. & Stuhlmüller, A. The design and implementation of probabilistic
561 programming languages. (2014).