Published as:
Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. Nature Human Behaviour. https://doi.org/10.1038/s41562-021-01145-1

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

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

[1] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology

[2] Department of Psychology, Stanford University

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

[*] These authors contributed equally to this work

Abstract

Before formal education begins, children typically acquire a vocabulary of thousands of words. This learning process requires the use of many different information sources in their social environment, including including their current state of knowledge and the context in which they hear words used. How is this information integrated? We specify a developmental model according to which children consider information sources in an age-specific way 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 in which they had to integrate three information sources. Model comparison suggests that the central locus of development is an increased sensitivity to individual information sources, rather than changes in integration ability. This work presents a developmental theory of information integration during language learning and illustrates how formal models can be used to make a quantitative test of the predictive and explanatory power of competing theories.

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

Published as:

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. Nature Human Behaviour. https://doi.org/10.1038/s41562-021-01145-1

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

Human communicative abilities are unrivaled in the animal kingdom.[1–3] Language – in whatever modality – is the medium that allows humans to collaborate and coordinate in species-unique ways, making it the bedrock of human culture and society.[4] Thus, to absorb the culture around them and become functioning members of society, children need to learn language.[5] A central problem in language learning is referent identification: To acquire the conventional symbolic relation between a word and an object, a child must determine the intended referent of the word. There is no unique cue to reference, however, that can be used across all situations.[6] Instead, referents can only be identified inferentially by reasoning about the speaker's intentions.[7–10] That is, the child has to infer what the speaker is communicating about based on information sources in the utterance's social context.

From early in development, children use several different mechanisms to harness social-contextual information sources.[7,9,11] Children expect speakers to use novel words for unknown objects,[12–15] to talk about objects that are relevant,[16,17] new in context,[18,19] or related to the ongoing conversation.[20–22] These different mechanisms, however, have been mainly described and theorized about in isolation. The implied picture of the learning process is that of a "bag of tricks": mechanisms that operate (and develop) independently from one another.[11] As such, this view of the learning process does not address the complexity of natural social interaction during which many sources of information are present.[6,23] How do children arbitrate between these sources in order to accurately infer a speaker's intention?

When information integration is studied directly, the focus is mostly on how children interpret or learn words in light of social-contextual information.[24–32] In one classic study,[33] 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 three of the four 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 speaker's visual perspective to correctly infer which of the glasses the speaker was referring to. This study advanced our understanding by documenting that preschoolers use both information sources, a finding confirmed by a variety of other work.[26,29,31] Yet these studies neither specify nor test the process by which children integrate different information sources. When interpreting such findings, work in this tradition refers to social-pragmatic theories of language use and learning,[9,10,34–36] 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,[9,10] our theory is based on the following premises: information sources serve different functional roles but are combined as part of an integrated social inference process.[34–37] 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[38–40]. In contrast to earlier modelling work, we treat word learning as the outcome of a social inference process instead of a cross-situational[41,42] or principle-based learning process.[43] 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.[44,45] First, we use the model to make quantitative predictions about children's behavior in new situations – predictions we test against new data. This form of model testing has been successfully used with adults[38,46] and here we extend it to children. Next, we quantify 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*. Alternative models either assume that children ignore some information sources or – in line with a "bag of tricks" approach – assume that children compute isolated inferences and then weigh their outcome in a post-hoc manner.

We focus on three information sources that play a central part in theorizing about language use and learning: (1) expectations that speakers communicate in a cooperative and informative manner,[12,16,35] (2) shared common ground about what is being talked about in conversation,[36,47,48] and (3) semantic knowledge about previously learned word–object mappings.[11,49]

Our *rational integration model* arbitrates between information sources via Bayesian inference (see Fig. 1f for model formulae). A listener (L_1) reasons about the referent of a speaker's (S_1) utterance. This reasoning is contextualized by the prior probability ρ of each referent. We treat ρ as a conversational prior which originates from the common ground shared between the listener and the speaker. This interpretation follows from the social nature of our experiments (see below). From a modelling perspective, ρ can be (and in fact has been) used to capture non-social aspects of a referent, for example its visual salience[38]. 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 θ_j (for object j) the speaker ascribes to the (literal) listener.

Taken together, this model provides a quantitative theory of information integration during language learning. The three information sources operate on different timescales: speaker informativeness is a momentary expectation about a particular utterance, common ground grows over the course of a conversation, and semantic knowledge is learned across development. This interplay of timescales has been hypothesized to be an important component of word meaning inference,[42,50] and we link these different time-dependent processes together via their proposed impact on model components. 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.[7,51] In the model, this is accomplished by creating age-dependent parameters capturing developmental changes in 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).

To test the predictive and explanatory power of our model, we designed a word-learning experiment in which we jointly manipulated the three information sources (Fig. 1). Children interacted on a tablet computer with a series of storybook speakers.[52] 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). The speakers used a novel word (e.g., "wug") in the context of two potential referents, and then the child was asked to identify a new instance of the novel word, testing their inference about the speaker's intended referent. To vary the strength of the child's inference, we systematically manipulated the familiarity of the known object (from e.g., the highly familiar "duck" to the relatively unfamiliar "pawn") and whether the familiar or novel object was new to the speaker (i.e., whether it was not part of common ground).

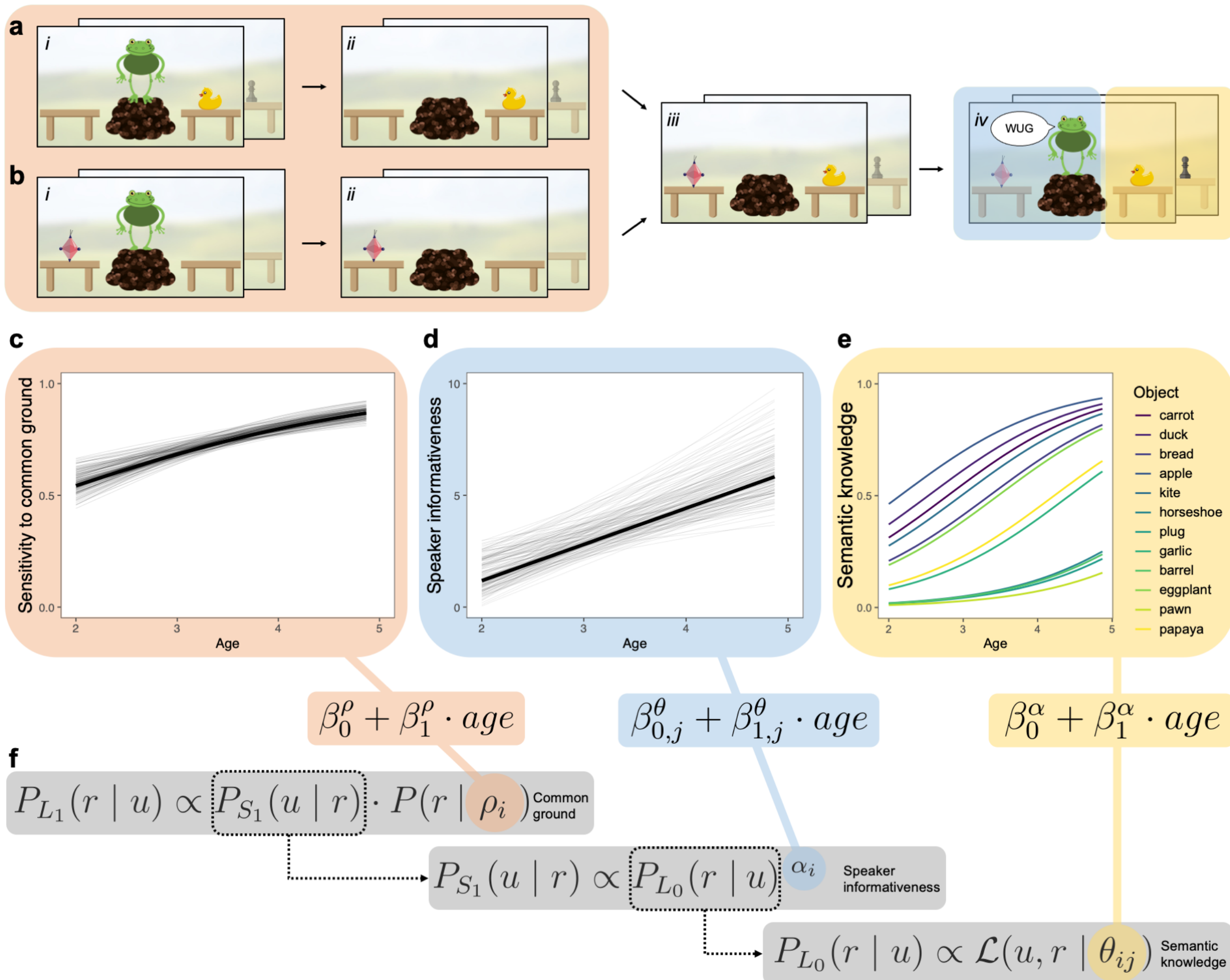

Figure 1. **Experimental task and model**. (a and b) Screenshots from the experimental task. (i) The speaker encounters one object and then leaves the scene. (ii) While the speaker is away, (iii) a second object appears, (iv) when returning, the speaker uses a novel word to request an object. Sections (i) to (iii) establish common ground between the speaker and the listener, in that one object is new in context (red). The request in (iv) licenses an inference based on expectations about how informative speakers are (blue). 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 – yellow). (a) shows the condition of the experiment 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. Developmental trajectories are shown for (c) sensitivity to common ground, (d) speaker informativeness and (e) semantic knowledge, estimated based on separate experiments (see main text). (f) gives the model equation for the rational integration model and links information sources to model parameters.

This paradigm allows us to examine the integration of the three information sources described above. First, the child may infer that a cooperative and informative speaker[12,16] 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 inference is often referred to as a "mutual exclusivity" inference.[13,15] Second, the child may draw 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.[18,19] 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 use 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.[53–55] Taken together, the child has the opportunity to integrate their assumptions about (1) cooperative communication, (2) their understanding of the common ground, and (3) their existing semantic knowledge. In one condition of the experiment, information sources were aligned (Fig. 1a) while in the other they were in conflict (Fig. 1b).

Results

Predicting information integration across development

We tested the model in its ability to predict 2–to–5 year-old children's judgments about word meaning. We estimated children's (N=148) developing sensitivities to individual information sources in two separate experiments (Experiments 1 and 2; see Fig. 1c-e). In Experiment 1, we estimated children's sensitivity to informativeness jointly with their semantic knowledge. In Experiment 2, we estimated sensitivity to common ground. We then generated parameter-free *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 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). All procedures, sample sizes, and analyses were pre-registered (see methods).

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). With predictions and data binned by child age (in years), the model explained 79% of the variance in the data (Fig. 2a). These results support the assumption of the model that children integrate all three available information sources.

It is still possible, however, that simpler models might make equally good – or even better – predictions. For example, work on children's use of statistical information during morphology learning showed that children's behaviour was best explained by a model which selectively ignored parts of the input.[56]

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 resulting *no word knowledge model* uses the same model architecture as the rational integration model. It uses 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 inferences do not vary depending on the particular familiar object but only on the age-specific average semantic knowledge (i.e., a marker of gross vocabulary size). 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 utterance. As a consequence, the inference is solely based on common ground expectations.

We found little support for these heuristic models (Fig. 2b). When using Bayesian model comparison via marginal likelihood of the data,[57] the data were several orders of magnitude more likely under the rational integration model compared to any of the lesioned models (Fig. 2). Figure 2c exemplifies the differences between the models: all heuristic models systematically underestimated children's performance in the congruent condition. Thus, even when the information sources are redundant (i.e., they all point to the same referent), children's inferences are notably strengthened by each of them. In the incongruent condition, the no word knowledge model underestimates performance, because it does not differentiate between the different familiar objects, and in the case of a highly familiar word such as "duck", underestimated the strength of the mutual exclusivity inference and its compensatory effect. The no speaker informativeness model completely ignores this inferences, which leads to even worse predictions. In contrast to the lesioned models that underestimated performance, the no common ground model overestimated performance in the incongruent condition because it ignored the dampening effect of common ground favoring a different referent. Taken together, we conclude that children considered all available information sources.

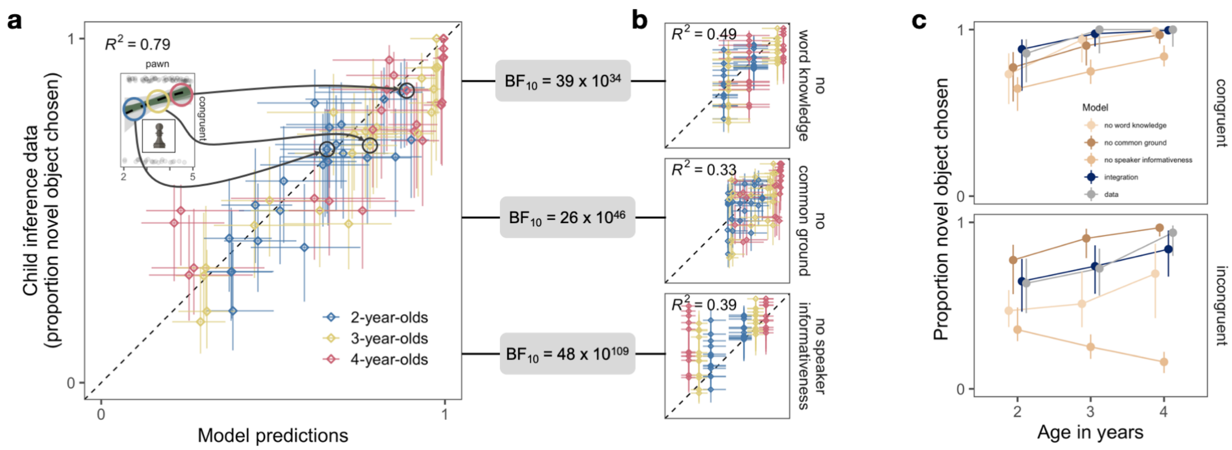

Figure 2. **Predicting information integration**. Correlation between model predictions and child inference data for all 24 conditions and for each age group (binned by year) for the rational integration model (a) and the three lesioned models (b). Horizontal and vertical error bars show 95% HDI. Inset shows an example of model predictions as developmental trajectories (see Fig. 3). BF_{10} gives the Bayes Factor in favor of the integration model based on the marginal likelihood of the data under each model. (c) Predictions from all models considered alongside the data (with 95% HDI) for two experimental conditions (familiar word: *duck*).

Explaining the process of information integration

In the previous section, we established that children integrated all available information sources to infer the meanings of new words. This result, however, does not speak to the process by which information is assumed to be integrated. Thus, in this section, we ask which integration process best explains children's behavior.

The rational integration model assumes that all information sources enter into a joint inference process, but alternative integration processes are conceivable and might be consistent with the data. For example, the "bag of tricks"[11] hypothesis mentioned in the introduction could be restated as a modular integration process: children might compute independent inferences based on subsets of the available information and then integrate them in a post-hoc manner by weighting them according to some parameter. 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)[12,13,53] while common ground information enters into a second inference. The outcomes of both processes are then weighted according to a bias parameter ϕ. Like the rational integration model, this model takes in all available information sources in an age-sensitive way and assumes that they are integrated. The only difference lies in the nature of the integration process: the biased integration model privileges some information sources over others in an ad-hoc manner.

The parameter ϕ in the biased integration model is unknown ahead of time and has to be estimated based on the experimental data. That is, through Experiments 1 and 2 alone, we do not learn anything about the relative importance of the information sources. As a consequence – and in contrast to the rational integration model – the biased integration model does not allow us to make *a priori* predictions about the new data (Experiment 3) in the way we described above. For a fair comparison, we therefore constrained the parameters in the rational integration model by the data from Experiment 3 as well. As a consequence, both models estimate their parameters using all the data available in a fully Bayesian manner (see Fig. S4).

The biased integration model makes reasonable posterior predictions and explains 78% of the variance in the data (Fig. 3b). The parameter ϕ – indicating the bias to one of the inferences – was estimated to favor the mutual exclusivity inference (Maximum A-Posteriori estimate = 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 the marginal likelihood of the data (Fig. 3). When constrained by the data from all experiments, the rational integration model explains 87% of the variance in the data. Fig. 3e exemplifies the difference between the models: the biased integration model puts extra weight on the mutual exclusivity inference and thus fails to capture performance when this inference is weak compared to the common ground inference – such as in the congruent condition for younger children. As a result, a fully integrated – as opposed to a modular and biased – integration process explained the data better.

The rational integration model assumes that the integration process itself does not change with age.[7] That is, while children's sensitivity to each information source develops, the way the information sources relate to one another remains the same. The biased integration model can provide the basis for an alternative proposal about developmental change, one in which the integration process itself changes with age. That is, children may be biased towards some information sources, and that bias itself may change 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 also explains a substantial portion of the variance in the data: 78% (Fig. 3c). The estimated developmental trajectory for the bias parameter ϕ suggests that younger children put a stronger 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). Looking at Figure 3e, we can see that the developmental bias model tends to underestimate children's performance because the supportive interplay between the different inferences is constrained. In the biased models, the overall inference can only be as strong as the strongest of the components – in the rational integration model, the components interact with one another, allowing for a stronger inference than the individual parts would suggest.

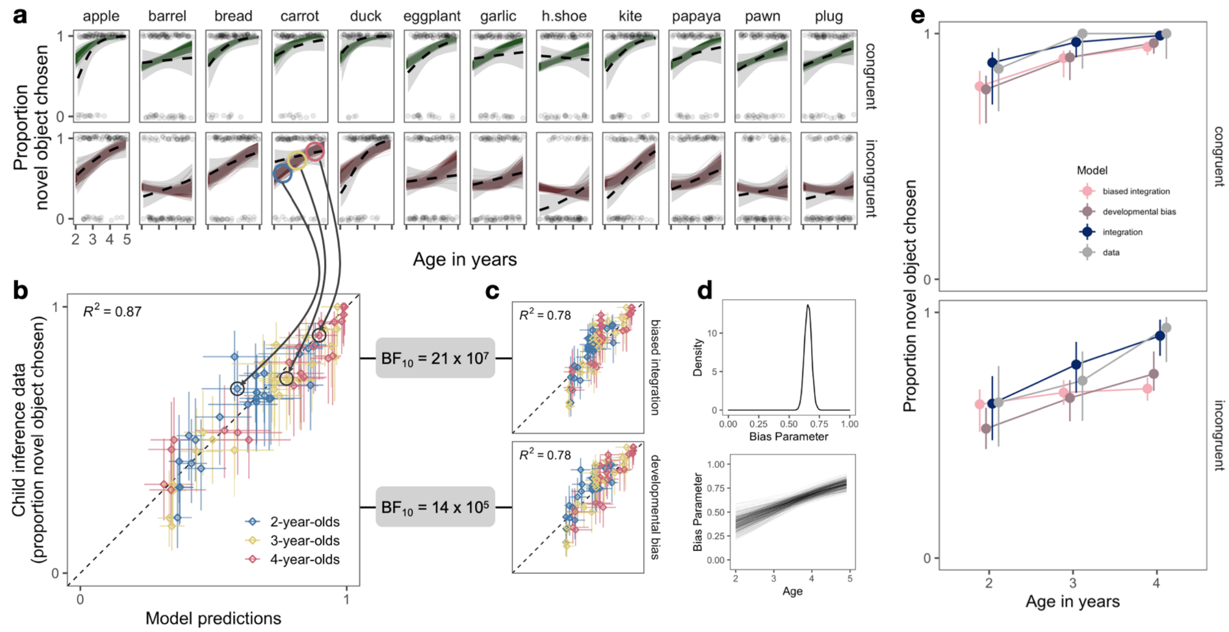

Figure 3. **Explaining information integration across development**. (a) Model predictions from the rational integration model (colored lines) next to the behavioral data (dotted black lines with 95% CI in gray) for all 24 experimental conditions. Top row (blue) shows congruent conditions, bottom row (red) shows incongruent conditions. Familiar 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 biased 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 (e) Predictions from all models considered alongside the data (with 95% HDI) for two experimental conditions (familiar word: duck).

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 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.

Previous work on the study of information integration during language comprehension focused on how adults combine perceptual, semantic or syntactic information.[58–62] Our work extends this work to the development of pragmatics. Our model is based on classic social-pragmatic theories on language use and comprehension.[10,34–36] As a consequence, instead of assuming that different information sources feed into separate word-learning mechanisms (the "bag of tricks" view), we assume that all of these information sources play a functional role in an integrated social inference process. Our model goes beyond previous theoretical and empirical work by describing the computations that underlie this inference process. Furthermore, we presented 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 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.

The rational integration model made accurate quantitative predictions across a range of experimental conditions both when information sources were aligned and when 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. We also 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 weighing the outcomes of these separate inferences by some parameter. 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 changes with age. In both cases, the rational integration model provided a much better fit to the data, suggesting that the integration process remains stable over time. That is, there is developmental continuity and therefore no qualitative difference in how a 2-year-old integrates information compared to a 4-year-old.

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.[38,39,63–67] 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, non-verbal aspects of the utterance (e.g. eye-gaze or gestures) can affect children's mutual exclusivity inferences.[68–72] As a first step in this direction, we recently studied how adults and children integrate non-verbal utterances with common ground[51]. Using a structurally similar rational integration model, we 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.[28,73,74]

Our model and empirical paradigm provide a foundation on which to test deeper questions about language development. First, our experiments should be performed in children from different cultural backgrounds, learning different languages.[75] In such studies, we would not expect our results to replicate in a strict sense; that is, we would not expect to see the same developmental trajectories in all cultures and languages. Substantial variation is much more likely. Studies on children's pragmatic inferences in different cultures have documented both similar[76,77] and different[78] developmental trajectories. Nevertheless, our model provides a way to think about how to reconcile cross-cultural variation with a shared cognitive architecture: We predict differences in how sensitive children are to the individual information sources at different ages, but similarities in how information is integrated.[7] In computational terms, we assume a universal architecture that specifies the relation between a set of varying parameters. Of course, either confirmation or disconfirmation of this prediction would be informative.

Second, it would be useful to flesh out the cognitive processes that underlie reasoning about common ground. The basic assumption that common ground changes interlocutors' expectations about what are likely referents[79] has been used in earlier modelling work on the role of common ground in reference resolution.[62] Here we went one step further and measured the strength of these expectations to inform the parameter values in our model. However, in its current form, our model treats common ground as a conversational prior and does not specify how the listener arrives at the expectation that some objects are more likely referents because they are new in common ground. That is, computationally, our model does not differentiate between common ground information and other reasons that might make an object contextually more salient. An interesting starting point to overcome this shortcoming would be modelling work on the role of common ground in conversational turn taking.[80]

Finally, our model is a model of referent identification in the moment of the utterance. At the same time, the constructs made use of by our model are shaped by factors that unfold across multiple time points and contexts: Common ground is built over the course of a conversation, and the lexical knowledge of a child is shaped across a language developmental time-scale. Even speaker informativeness could be imagined to vary over time following repeated interactions with a particular speaker. What is more, assessing speaker informativeness is unlikely to be the outcome of a single, easy-to-define process. The expectations about informative communication that we take it to represent are probably the result of the interplay between multiple social and non-social inference processes. The broader point here is that, our model makes use of unidimensional representations of high-dimensional, structured processes and examines how these representations are integrated. As such, it is first and foremost a computational description of the inferences and we therefore make no strong claims about the psychological reality of the parameters in it. Connecting our model with other frameworks that focus on the cognitive, temporal and cross-situational aspects of word learning would elucidate further these complex processes.[42,50,81]

This work advances our understanding of how children navigate the complexity of their learning environment. Methodologically, it illustrates how computational models can be used to test theories; 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 sizes, and analysis for each experiment were pre-registered (https://osf.io/7rg9j/registrations; dates of registration: 05/02/2019, 04/05/2019, 03/02/2019). Experimental procedures, data, model and analysis scripts can be found in an online repository (https://github.com/manuelbohn/spin). Experiments 1 and 2 were designed to estimate children's developing sensitivity to each information source. The results of these experiments determine the parameter values in the model (see Fig. 1 c-f). Experiment 3 was designed to test how children integrate different information sources.

Participants

Sample sizes for each experiment were chosen to have at least 30 data points per cell (i.e. unique combination of condition, item and age-group). 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 were 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 were 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 were 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 in a children's museum in San José, California, USA. This population is characterized by a 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. Tablet-based storybooks are commonly used to simulate social interactions in developmental research and interventions.[82] A recent, direct comparison found similar performance with tablet-based and printed storybooks in a word learning paradigm.[52] Furthermore, our results in Experiment 1 and 2 replicate earlier studies on mutual exclusivity and discourse novelty that used live interactions instead of storybooks.[18,19]

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 mutual exclusivity inference.[13,53] On one table, there was a familiar object, on the other table, there was an unfamiliar object (a novel design 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 a response as a 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. We used familiar objects that we expected to vary along the dimension of how likely children were to know the word for it. This set 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 [chess piece]). The selection was based on the age of acquisition ratings from Kuperman and colleagues.[83] 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 built 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.[18,19] 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 at the beginning of the experiment, the speaker as well as the location the speaker turned to first was counterbalanced. Children completed ten trials, each with a different pair of unfamiliar objects. We coded a response as a correct choice if children chose as the referent of the novel word the object that was new to the speaker.

Experiment 3 combined the procedures from Experiments 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 ("Oh cool, there is a [non-word] on the table, how neat, can you give me the [non-word]?"). 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 (i.e., it was both novel to the speaker in the context and it was an object that does not have a label in the lexicon). 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 alignments (congruent vs incongruent) x 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 pre-registered 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 cognitive models and Bayesian data analytic models were implemented in the probabilistic programming language `WebPPL`.[84] The corresponding model code can be found in the associated online repository. Information about priors for parameter estimation and Markov chain Monte Carlo settings can also be found in the supplementary information and the online repository.

As a first step, we used the data from Experiments 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 model 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 semantic knowledge and speaker informativeness parameters in an age-sensitive manner. Specifically, we inferred the intercepts and slopes for speaker informativeness via a linear regression submodel and semantic knowledge via a logistic regression submodel, the values of which were then combined in the cognitive model to generate model predictions to predict 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 "prediction" models, the parameters whose values were inferred by the data from Experiments 1 & 2 were then used to make out-of-sample predictions for Experiment 3. 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 a mapping from an utterance u to a referent r, defined as

$$P_{L_1}^{int}(r \mid u; \{\rho_i, \alpha_i\, \theta_{ij}\}) \propto P_{S_1}(u \mid r; \{\alpha_i, \theta_{ij}\}) \cdot P(r \mid \rho_i) \tag{1}$$

where i represents the age of the participant and the j the familiar object. 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:

$$P_{L_1}^{no\_wk}(r \mid u; \{\rho_i, \alpha_i\, \theta_i\}) \propto P_{S_1}(u \mid r; \{\alpha_i, \theta_i\}) \cdot P(r \mid \rho_i) \tag{2}$$

and the only difference lies in the parameter θ, which does not vary as a function of j, the object (i.e., θ in this model is analogous to a measure of gross vocabulary development). 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

$$P_{L_1}^{no\_cg}(r \mid u; \{\alpha_i\, \theta_{ij}\}) \propto P_{S_1}(u \mid r; \{\alpha_i, \theta_{ij}\}) \tag{3}$$

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

$$P_{L_1}^{no\_si}(r \mid u; \{\rho_i\}) \propto P(r \mid \rho_i) \tag{4}$$

As noted above, the explanation models used parameters that were additionally constrained by the data from Experiment 3, but the way these parameters were combined in the rational integration model was the same as above. The biased integration model is defined as

$$P_{L_1}^{biased}(r \mid u; \{\phi, \rho_i, \alpha_i, \theta_{ij}\}) = \phi \cdot P_{ME}(r \mid u; \{\alpha_i, \theta_{ij}\}) + (1 - \phi) \cdot P(r \mid \rho_i) \tag{5}$$

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

$$P_{L_1}^{dev\_bias}(r \mid u; \{\phi_i, \rho_i, \alpha_i, \theta_{ij}\}) = \phi_i \cdot P_{ME}(r \mid u; \{\alpha_i, \theta_{ij}\}) + (1 - \phi_i) \tag{6}$$

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 distribution on parameters; the pairwise ratio of marginal likelihoods for two models is known as the Bayes Factor. It is interpreted as how many times more likely the data is under one model compared to the other. Bayes Factors quantify the quality of predictions of a model, averaging over the possible values of the parameters of the models (weighted by the prior probabilities of those parameter values); by averaging over the prior distribution on parameters, Bayes Factors implicitly take into account model complexity because models with more parameters will tend to have a broader prior distribution over parameters, which in effect, can water down the potential gains in predictive accuracy that a model with more parameters can achieve.[57] For this analysis, we treated age continuously.

References

1.

Taylor, C. *The language animal.* (Harvard University Press, 2016).

2.

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

3.

Fitch, W. T. *The evolution of language.* (Cambridge University Press, 2010).

4.

Smith, J. M. & Szathmary, E. *The major transitions in evolution.* (Oxford University Press, 1997).

5.

Tomasello, M. *A natural history of human thinking.* (Harvard University Press, 2018).

6.

Frank, M. C., Tenenbaum, J. B. & Fernald, A. Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development* **9**, 1–24 (2013).

7.

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

8.

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

9.

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

10.

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

11.

Bloom, P. *How children learn the meanings of words.* (MIT Press, 2002).

12.

Clark, E. V. The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition* (1987).

13.

Markman, E. M. & Wachtel, G. F. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology* **20**, 121–157 (1988).

14.

Carey, S. & Bartlett, E. Acquiring a single new word. *Proceedings of the Stanford Child Language Conference* **15**, 17–29 (1978).

15.

Halberda, J. The development of a word-learning strategy. *Cognition* **87**, B23–B34 (2003).

16.

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

17.

Schulze, C., Grassmann, S. & Tomasello, M. 3-year-old children make relevance inferences in indirect verbal communication. *Child Development* **84**, 2079–2093 (2013).

18.

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

19.

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).

20.

Bohn, M., Le, K., Peloquin, B., Koymen, B. & Frank, M. C. Children's interpretation of ambiguous pronouns based on prior discourse. *Developmental Science* **e13049**, (2020).

21.

Sullivan, J. & Barner, D. Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. *Developmental Science* **19**, 63–75 (2016).

22.

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).

23.

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).

24.

Hollich, G. J., Hirsh-Pasek, K. & Golinkoff, R. M. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development* **65**, i–135 (2000).

25.

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

26.

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

27.

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

28.

Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B. & Schulz, L. E. Social pragmatics: Preschoolers rely on commonsense psychology to resolve referential underspecification. *Child Development* **91**, 1135–1149 (2020).

29.

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

30.

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

31.

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

32.

Kachel, G., Hardecker, D. J. K. & Bohn, M. Young children's developing ability to integrate gestural and emotional cues. *Journal of Experimental Child Psychology* **201**, 104984 (2021).

33.

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

34.

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

35.

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

36.

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

37.

Shafto, P., Goodman, N. D. & Frank, M. C. Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science* **7**, 341–351 (2012).

38.

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

39.

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

40.

Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review* **26**, 1596–1618 (2019).

41.

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

42.

Frank, M. C., Goodman, N. D. & Tenenbaum, J. B. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* **20**, 578–585 (2009).

43.

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

44.

Shmueli, G. To explain or to predict? *Statistical Science* **25**, 289–310 (2010).

45.

Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science* **12**, 1100–1122 (2017).

46.

Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).

47.

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

48.

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

49.

Fenson, L. *et al.* Variability in early communicative development. *Monographs of the Society for Research in Child Development* i–185 (1994).

50.

McMurray, B., Horst, J. S. & Samuelson, L. K. Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review* **119**, 831 (2012).

51.

Bohn, M., Tessler, M. H., Merrick, M. & Frank, M. C. Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings. (2019). doi:10.31234/osf.io/xma4f

52.

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

53.

Lewis, M. L., Cristiano, V., Lake, B. M., Kwan, T. & Frank, M. C. The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition* **198**, 104191 (2020).

54.

Grassmann, S., Schulze, C. & Tomasello, M. Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology* **6**, 1200 (2015).

55.

Ohmer, X., König, P. & Franke, M. Reinforcement of semantic representations in pragmatic agents leads to the emergence of a mutual exclusivity bias. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*

56.

Gagliardi, A., Feldman, N. H. & Lidz, J. Modeling statistical insensitivity: Sources of suboptimal behavior. *Cognitive Science* **41**, 188–217 (2017).

57.

Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: A practical course.* (Cambridge University Press, 2014).

58.

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

59.

Kamide, Y., Scheepers, C. & Altmann, G. T. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research* **32**, 37–55 (2003).

60.

Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. Integration of word meaning and world knowledge in language comprehension. *Science* **304**, 438–441 (2004).

61.

Özyürek, A., Willems, R. M., Kita, S. & Hagoort, P. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* **19**, 605–616 (2007).

62.

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

63.

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

64.

Wang, S., Liang, P. & Manning, C. D. Learning language games through interaction. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* 2368–2378 (2016).

65.

Tessler, M. H. & Goodman, N. D. The language of generalization. *Psychological Review* **126**, 395–436 (2019).

66.

Yoon, E. J., Tessler, M. H., Goodman, N. D. & Frank, M. C. Polite speech emerges from competing social goals. *Open Mind* **4**, 71–87 (2020).

67.

Savinelli, K., Scontras, G. & Pearl, L. Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (2017).

68.

Paulus, M. & Fikkert, P. Conflicting social cues: Fourteen-and 24-month-old infants' reliance on gaze and pointing cues in word learning. *Journal of Cognition and Development* **15**, 43–59 (2014).

69.

Graham, S. A., Nilsen, E. S., Collins, S. & Olineck, K. The role of gaze direction and mutual exclusivity in guiding 24-month-olds' word mappings. *British Journal of Developmental Psychology* **28**, 449–465 (2010).

70.

Gangopadhyay, I. & Kaushanskaya, M. The role of speaker eye gaze and mutual exclusivity in novel word learning by monolingual and bilingual children. *Journal of Experimental Child Psychology* **197**, 104878 (2020).

71.

Jaswal, V. K. & Hansen, M. B. Learning words: Children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science* **9**, 158–165 (2006).

72.

Grassmann, S. & Tomasello, M. Young children follow pointing over words in interpreting acts of reference. *Developmental Science* **13**, 252–263 (2010).

73.

Bridgers, S., Jara-Ettinger, J. & Gweon, H. Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour* **4**, 144–152 (2020).

74.

Jara-Ettinger, J., Gweon, H., Schulz, L. E. & Tenenbaum, J. B. The naive utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences* **20**, 589–604 (2016).

75.

Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology* **162**, 31–38 (2017).

76.

Su, Y. E. & Su, L.-Y. Interpretation of logical words in mandarin-speaking children with autism spectrum disorders: Uncovering knowledge of semantics and pragmatics. *Journal of Autism and Developmental Disorders* **45**, 1938–1950 (2015).

77.

Zhao, S., Ren, J., Frank, M. C. & Zhou, P. The development of quantity implicatures in mandarin-speaking children. (2019).

78.

Fortier, M., Kellier, D., Flecha, M. F. & Frank, M. C. Ad-hoc pragmatic implicatures among shipibo-konibo children in the peruvian amazon. *psyarxiv.com/x7ad9* (2018). doi:10.31234/osf.io/x7ad9

79.

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

80.

Anderson, C. J. Tell me everything you know: A conversation update system for the rational speech acts framework. *Proceedings of the Society for Computation in Linguistics* **4**, 244–253 (2021).

81.

Yurovsky, D. & Frank, M. C. An integrative account of constraints on cross-situational learning. *Cognition* **145**, 53–62 (2015).

82.

Richter, A. & Courage, M. L. Comparing electronic and paper storybooks for preschoolers: Attention, engagement, and recall. *Journal of Applied Developmental Psychology* **48**, 92–102 (2017).

83.

Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods* **44**, 978–990 (2012).

84.

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

Acknowledgements

M. Bohn received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 749229. M. H. Tessler was funded by the National Science Foundation SBE Postdoctoral Research Fellowship Grant No. 1911790. M. C. Frank was supported by a Jacobs Foundation Advanced Research Fellowship and the Zhou Fund for Language and Cognition.The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

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