The role of experience in disambiguation during early word learning

Molly Lewis[1, 6], Veronica Cristiano[2], Brenden Lake[3], Tammy Kwan[4], & Michael C. Frank[5]

[1] University of Chicago

[2] Gallaudet University

[3] New York University

[4] Cognitive Toybox, Inc.

[5] Stanford University

[6] University of Wisconsin-Madison

Author Note

Data from Experiment 2 were previously presented in the Proceedings of the Cognitive Science Society Conference in Lewis & Frank (2013). *To whom correspondence should be addressed. E-mail: mollylewis@uchicago.edu

Abstract

Young children tend to map novel words to novel objects even in the presence of familiar competitors, a finding that has been dubbed the "disambiguation" effect. Theoretical accounts ofthis effect have debated whether it is due to initial constraints on children's lexicons (e.g. a principle of mutual exclusivity) or situation-specific pragmatic inferences. We present synthesis of exisiting evidence on this phenomonon through a meta-analysis of the existing literature. We then present two experiments that help distinguish between these theoretical constraints. We conclude by suggesting that multiple cognitive mechanisms may underlie the disambiguation effect in word learning.

Keywords: keywords

Word count: X

The role of experience in disambiguation during early word learning

Introduction

A central property of language is that each word in the lexicon maps to a unique concept, and each concept maps to a unique word (Clark, 1987). Like other important regularities in language (e.g., grammatical categories), children cannot directly observe this general property. Instead, they must learn to use language in a way that is consistent with the generalization on the basis of evidence about only specific word-object pairs.

Even very young children behave in a way that is consistent with this one-to-one regularity in language. Evidence for this claim comes from what is known as the "disambiguation" or "mutual exclusivity" (ME) effect (we return to the issue of nomenclaturer below). In a typical demonstration of this effect (Markman & Wachtel, 1988), children are presented with a novel and familiar object (e.g., a whisk and a ball), and are asked to identify the referent of a novel word ("Show me the dax"). Children in this task tend to choose the novel object as the referent, behaving in a way that is consistent with the one-to-one word-concept regularity in language across a wide range of ages and experimental paradigms (Bion, Borovsky, & Fernald, 2012; Golinkoff, Mervis, Hirsh-Pasek, & others, 1994; J. Halberda, 2003; Markman, Wasow, & Hansen, 2003; Mervis, Golinkoff, & Bertrand, 1994).

This effect has received much attention in the word learning literature because the ability to identify the meaning of a word in ambiguous contexts is, in essence, the core problem of word learning. That is, given any referential context, the meaning of a word is underdetermined (Quine, 1960), and the challenge for the world learner is to identify the referent of the word within this ambiguous context. Critically, the ability to infer that a novel word maps to a novel object makes the problem much easier to solve. For example, suppose a child hears the novel word "kumquat" while in the produce aisle of the grocery store. There are an infinite number of possible meanings of this word given this referential context, but the child's ability to correctly disambiguate would lead her to rule out all meanings for which she already had a name. With this restricted hypothesis space, the child is more likely to identify the correct referent than if all objects in the context were considered as possible referents.

Despite – or perhaps due to – the attention that the ME effect has received, there is little consensus regarding the cognitive mechanisms underlying it. Does it stem from a basic inductive bias on children's learning abilities ("bias accounts," see below), a learned regularity about the structure of language ("overhypothesis accounts"), reasoning about the goals of communication in context ("pragmatic accounts"), or perhaps some mixture of these? The goal of the current manuscript is to lay out these possibilities and discuss the state of the evidence. Along the way we present a meta-analysis of the extant empirical literature. We then present two new, relatively large-sample developmental experiments that investigate the dependence of children's ME inferences on vocabulary (Experiment 1) and experience with particular words (Experiment 2). We end by discussing the emergence of ME inferences in a range of computational models of word learning. We conclude that:

1. Explanations of ME are not themselves mutually exclusive and likely more than one is at play;
2. The balance of responsibility for behavior likely changes developmentally, with basic biases playing a greater role for younger children and learned overhypotheses playing a greater role for older children.
3. All existing accounts put too little emphasis on the role of experience and strength of representation; this lack of explicit theory in many cases precludes definitive tests.
4. ME inferences are distinct from learning.

A note on terminology.

Markman and Wachtel (1988)'s seminal paper coined the term "mutual exclusivity," which was meant to label the theoretical proposal that "children constrain word meanings by assuming at first that words are mutually exclusive – that each object will have one and only one label." (Markman, 1990, p. 66). That initial paper also adopted a task used by a variety of previous authors (CHECK THESE CITES, including **???**, **???**, **???**), in which a novel and a familiar object were presented to children in a pair and the child was asked to "show me the *x*," where *x* was a novel label. Since then, informal discussions have used the same name for the paradigm and effect (selecting the novel object as the referebnt of the novel word) as well as the theoretical account (an early assumption or bias). This conflation of paradigm/effect with theory is problematic, as other authors who have argued against the theoretical account then are in the awkward position of rejecting the name for the paradigm they have used. Other labels (e.g. "disambiguation" or "referent selection" effect) are not ideal, however, because they are not as specific do not refer as closely to the previous literature. Here we adopt the label "mutual exclusivity" (ME) for the general family of paradigms and associated effects, *without* prejudgment of the theoretical account of these effects.

ME has also been referred to as "fast mapping." This conflation is confusing at best. In an early study, S. Carey and Bartlett (1978) presented children with an incidental word learning scenario by using a novel color term to refer to an object: "You see those two trays over there. Bring me the *chromium* one. Not the red one, the *chromium* one." Those data (and subsequent replications, e.g. L. Markson & Bloom, 1997) showed that this exposure was enough to establish some representation of the link between phonological form and meaning that endured over an extended period; a subsequent clarification of this theoretical claim emphasized that these initial meanings are partial (S. Carey, 2010). Importantly, however, demonstrations of retention relied on learning in a case where there was a contrastive presentation of the word with a larger set of contrastive cues (S. Carey & Bartlett, 1978) or pre-exposure to the object (L. Markson & Bloom, 1997).

Theoretical views of "mutual exclusivity"

What are the cognitive processes underlying this effect? A range of proposals in the literature.

Constraint and bias accounts. Under one proposal, Markman and colleagues (Markman & Wachtel, 1988, Markman et al. (2003)) suggest that children have a constraint on the types of lexicons considered when learning the meaning of a new word – a "mutual exclusivity constraint." With this constraint, children are biased to consider only those lexicons that have a one-to-one mapping between words and objects. Importantly, this constraint can be overcome in cases where it is incorrect (e.g. property names), but it nonetheless serves to restrict the set of lexicons initially entertained when learning the meaning of a novel word. Under this view, then, the disambiguation effect emerges from a general constraint on the structure of lexicons. This constraint is assumed to be innate or early emerging.

N3C

Probabilistic accounts. Regier

McMurray

Frank Goodman Tenenbaum

Fazly

Over-hypothesis accounts. Lewis & Frank (2013)

Pragmatic accounts. The disambiguation effect is argued to result from online inferences made within the referential context (Clark, 1987, Diesendruck and Markson (2001)). In particular, Clark suggests that the disambiguation effect is due to two pragmatic assumptions held by speakers. The first assumption is that speakers within the same speech community use the same words to refer to the same objects ("Principle of Conventionality"). The second assumption is that different linguistic forms refer to different meanings ("Principle of Contrast"). In the disambiguation task described above, then, children might reason (implicitly) as follows: You used a word I've never heard before. Since, presumably we both call a ball "ball" and if you'd meant the ball you would have said "ball," this new word must refer to the new object. Thus, under this account, the disambiguation effect emerges not from a higher-order constraint on the structure of lexicons, but instead from in-the-moment inferences using general pragmatic principles.

These two proposals have traditionally been viewed as competing explanations of the disambiguation effect. Research in this area has consequently focused on identifying empirical tests that can distinguish between these two theories. For example, Diesendruck and Markson (2001) compare performance on a disambiguation task when children are told a novel fact about an object relative to a novel referential label. They found that children disambiguated in both conditions and argued on grounds of parsimony that the same pragmatic mechanism was likely to be responsible for both inferences. More recent evidence contradicts this view: tests of children with autism, who are known to have impairments in pragmatic reasoning find comparable performance on the disambiguation task between typically developing children and children with autism (de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Preissler & Carey, 2005). This result provides some evidence for the view that disambiguation is due to a domain-specific lexical constraint.

Clark?

In the moment

Learned pragmatics

Logical inference accounts. Justin Halberda (2003)

Theory-constraining findings

NN vs. NF

Speaker-change studies

Autism

Bilingualism

Fast mapping + no retention

Developmental change (halberda)

Synthesis

These are definitely features of a successful account: Timescales - must be one “in the moment” - and one longer-term learned mechanism

Experience

Probabilistic representations

Could be the case also that it’s a mixture of pragmatic, etc.

We suggest this competing-alternatives approach to the disambiguation effect should be reconsidered. In a disambiguation task, learners may be making use of both general knowledge about how the lexicon is structured as well as information about the pragmatic or inferential structure of the task. Both of these constraints would then support children’s inferences. In other words, these two classes of theories may be describing distinct, complimentary mechanisms that each contribute to a single empirical phenomenon with their weights in any given task determined by children’s age and language experience, the nature of the pragmatic situation, and other task-specific factors.

The current study

Gather evidence on strength of finding

Test emergent relationship to vocabulary (E1)

Test causal relationship to representation strength (E2)

Re-evaluate

Meta-analysis

Methods

Search strategy. We conducted a forward search based on citations of Markman and Wachtel (1988) in Google Scholar, and by using the keyword combination “mutual exclusivity” in Google Scholar (September 2013; November 2017). Additional papers were identified through citations and by consulting experts in the field. We then narrowed our sample to the subset of studies that used one of two different paradigms: (a) an experimenter says a novel word in the context of a familiar object and a novel object and the child guesses the intended referent (the canonical paradigm; "Familiar-Novel"), or (b) experimenter first provides the child with an unambigous mapping of a novel label to a novel object, and then introduces a second novel object and asks the child to identify the referent of a second novel label ("Novel-Novel"). For Familiar-Novel conditions, we included conditions that included more than one familiar object (e.g. Familiar-Familar-Novel). From these conditions, we restricted our sample to only those that satisfied the following criteria: (a) participants were children (less than 12 years of age)[1], (b) referents were objects or pictures (not facts or object parts), and (c) no incongruent cues (e.g. eye gaze at familiar object). All papers used either forced-choice pointing or eye-tracking methodology. All papers were peer-reviewed with the exception of two dissertations (Williams, 2009; Frank, I., 1999), but all main results reported below remain the same when these papers are excluded. In total, we identified 43 papers that satisfied our selection criteria and had sufficient information to calculate an effect size.

Coding. For each paper, we coded separately each relevant condition with each age group entered as a separate condition. For each condition, we coded the paper metadata (citation) as well as several potential moderator variables: mean age of infants, method (pointing or eyetracking), participant population type, estimates of vocabulary size from the Words and Gestures form of the MacArthur-Bates Communicative Development Inventory when available (MCDI; **???**, Fenson et al. (2007)), referent type (object or picture), and number of alternatives in the forced choice task. We coded participant population as one of three subpopulationns that have studied in the literature: (a) typically-developing monolingual chilldren, (b) multilingual children (including both bilingual and trilingual children), and (c) non-typically developing children. Non-typically developing conditions included children with selective language imparement, language delays, hearing imparement,

[1]This cutoff was arbitrary but allowed us to include conditions from older children from non-typically-developing populations.

autism spectrum disorder, and down-syndrome.

In order to estimate effect size for each conditions, we also coded sample size, proportion novel-object selections, baseline (e.g., .5 in a 2-AFC paradigm), and standard deviations for novel object selections, t-statistic, and Cohen's d. For several conditions, there was insufficient data reported in the main text to calculate an effect size (no means and standard deviations, t-statistics, or Cohen's ds), but we were able to esimtate the means and standard deviations though measurement of plots ($N = 13$), imputation from other data within the paper ($N = 4$; see SI for details), or through contacting authors ($N = 26$). Our final sample included 157 effect sizes ($N_{\text{typical-developing}} = 135$; $N_{\text{multilingual}} = 12$; $N_{\text{non-typically-developing}} = 10$).

Statistical approach. We calculated effect sizes (Cohen's d) from reported means and standard deviations where available, otherwise we relied on reported test-statistics (t or d). Effect sizes were computed by a script, `compute_es.R`, available in the Github repository. All analyses were conducted with the metafor package (**???**) using mixed-effect models with grouping by paper.[2] In models with moderators, moderators variables were included as additive fixed effects. All estimate ranges are 95% confidence intervals.

Meta-analytic Analyses

We conducted a separate meta-analysis for four theoretically-relevant conditions: Familiar-Novel trials with typically developing participants, Novel-Novel trials with typically developing participants, conditions with multilingual participants, and conditions with non-typically developing participants.

Typically-Developing Population: Novel-Familiar Trials. We first examined effect sizes for the disambiguation effect for typically-developing children in the canonical familiar-novel paradigm. This is the central data point that theories of disambiguation must explain.

[2]The exact model specification was as follows: $model < -metafor :: rma.mv(yi = effect_size, V = effect_size_var, random = 1|paper, data = d)$.

Results. The overall effect size for these conditions was 1.1 [0.79, 1.42], and reliably greater than zero (p <.001). The effect sizes contained considerable heterogenity, however (Q = 968.13; p <.001).

We next tried to predict this heterogenity with two key moderators: age and vocabulary. In a model with age as a moderator, age was a reliable predictor of effect size (β = 0.05, z = 11.85, p <.001; see Table X), suggesting that the disambiguation effect becomes larger as children get older. Age of participants was highly correlated with vocabulary size in our sample ($r = 0.65$, $p < .01$), so next we asked whether vocabulary size predicted independent variance in the magnitude of the disambiguation bias on the subset of conditions for which we had estimates of vocabulary size ($N = 23$). To test this, we fit a model with both age and vocabulary size as moderators. Vocabulary size ($\beta = 0.07$, z = 2.14, $p = 0.03$), but not age (β = -0.78, z = -1.11, $p = 0.27$, was a reliable predictor of disambiguation effect size. ACTUALLY THIS ISN'T TRUE (true only for full model)

These analyses confirm that the disambiguation phenomenon is robust, and associated with a relatively large effect size (d = 1.1 [0.79, 1.42]). In addition, this set of analyses provides theory-constraining evidence about the mechansims underlying the effect. In particular, the finding that vocabulary predicts more variance in effect size, compared to age, suggests that there is an experience related component to the mechanism, independent of pure maturational development.

Typically-Developing Population: Novel-Novel Trials. The results from the Familiar-Novel trials point to a role for vocabulary knowledge in the strength of the disambiguation effect. One way in which this vocabulary knowledge could lead to increased performance on the Familiar-Novel disambiguation task is through increased certainty about the label associated with the familiar word: If a child is less certain that a ball is called "ball," then the child should be less certain that the novel label applies to the novel object. Novel-Novel trials control for potential variability in certainty about the familiar object by teaching participants a new label for a novel object prior to the critical disambiguation trial, where this previously-learned label becomes the "familar" object in the disambiguation trial. If knowledge of the familiar object is not the only contributor to age-related changes in the disambiguation effect, then there should be developmental change in Novel-Novel trials, as well as Novel-Familiar trials. In addition, if the strength of knowledge of the "familiar" object influences the strength of the disamiguation effect, then the overall effect size should be smaller for Novel-Novel trials, compared to Familar-Novel trials.

For conditions with the Novel-Novel trial design, the overall effect size was 1.36 [0.6, 2.11] and reliably greater than zero (p <.001). We next asked whether age predicted some of the variance in these trials by fitting a model with age as a moderator. Age was a reliable predictor of effect size (β = 0.03, z = 3.55, p <.001), suggesting that the strength of the disambiguation bias increases with age.

Finally, we fit a model with both age and trial type (Familar-Novel or Novel-Novel) as moderators of the disambiguation effect. Both moderators predicted independent variance in disambiguation effect size (age: β = -0.08, z = -0.42, p = 0.68; trial-type: β = 0.04, z = 12.34, p = 0), with Familiar-Novel conditions and conditions with older participants tending to have larger effect sizes.

These analyses point to an influence on the disambiguation effect of both development (either via maturation or experience-related changes) as well as the strength of the familiar word representation. A successful theory of disambiguation will need to account for both of these empirical facts.

Multilingual Population. We next turn to a different population of participants: Children who are simultaneously learning multiple languages. This popuiation is of theoretical interest because it allows us to isolate the influence of linguistic knowledge from the influence of domain-general capabilities. If the disambiguation phenomenon relies on mechanisms that are domain-general and independent of linguistic knowledge, then we should expect the magnitude of the effect size to be the same for multilingual children compared to monolingual children.

Children learning multiple languages reliably showed the disambiguation effect (d = 1.57 [0.69, 2.44]). We next fit a model with both monolingual (typically-developing) and multilingual participants, predicting effect size with language status (monolingual vs. multilingual), while controling for age. Language status was not a reliable predictor of effect size (β = 0.20, z = 1.42, p = 0.16), but age was (β = 0.03, z = 11.54, p = 0).

These data do not provide strong evidence that language-specific knoweldge influences effect size, however, the small sample size of studies from this population limit the power of this model to detect a difference if one existed.

Non-Typically-Developing Population. Finally, we examine a third-population of participants: non-typically developing children. This group includes a heterogenous sample of children with diagnoses including Autism-Spectrum Disorder (ASD), Mental Retardation, Williams Syndrome, Late-Talker, Selective Language Imparment, and deaf or hard-of-hearing participants. These populations are of theoretical interests because they allow us to observe how imparement to a particular aspect of cognition influences the magnitutude of the disambiguation effect. For example, children with ASD are thought to have impared social reasoning skills (e.g., Phillips, Baron-Cohen, & Rutter, 1998); thus, if children with ASD are able to succeed on disambiguation tasks, this suggests that social reasoning skills are not necessary to making a disambiguation inference.

Overall, non-typically developing children succeeded on disambiguation tasks (d = 1.57 [0.69, 2.44]). In a model with age as a moderator, age was a reliable predictor of the effect, suggesting children became more accurate with age, as with other populations (β = 0.04, z = 3.15, p <.001).

We also asked whether the effect size for non-typically developing children differed from typically-developing children, controling for age. We fit a model predicting effect size with both development type (typical vs. non-typical) and age. Development type was a reliable predictor of effect size with non-typically developing children tending to have a smaller bias compared to typically developing children (β = -0.50, z = -2.86, p = 0). Age was also a reliable predictor of effect size in this model ($\beta = 0.04$, $z = 11.34$, $p = 0$).

This analysis suggests that non-typically developing children succeed in the disambiguation paradigm just as typically developing children do, allbeit at lower rates. Theoretical accounts of the disambiguation phenomenon will need to account for how non-typically developing children are able to succeed in the disambiguation task, despite a range of different cognitive impairements.

Discussion

To summarize our meta-analytic findings, we find that there is a robust disambiguation effect across all four populations we studied and that, perhaps with the exception of mulitlinguals, the magnitude of this effect increases across development.

Taken together, these analyses provide several theoretical constraints about the mechanism underlying the disambiguation effect. First, language experience likely accounts for some developmental change. This conclusion derives from the fact that we see a larger effect size in Novel-Familiar trials compared to Novel-Novel trials, and that there is a suggestive correlation between vocabulary size and mutual exclusivity. Second, language experience is not sufficient to account for all developmental change in the effect. This constraint comes from the fact that we observe a larger bias with development in the Novel-Novel conditions, for which prior language experience is not relevant. In addition, there is no significant imparement to the disambiguation bias in multilingual children (who presumably have less language experience with any particular language), suggesting a role for domain-general abilities underlying the effect. Third, children with a range of different imparements are able to make the inference, suggesting that multiple mechanisms likely underly the effect across children.

These three constraints are consistent with many of proposed accounts individual, as well as a variety of combinations of them. In particular, an effect of language experience on the disambiguation effect via vocabulary knolwedge is consistent only with the overypothesis account, which predicts a stronger learned bias with vocabulary development. However, all four accounts are able to account for the developmental change in the NN trials. Under the overhypothesis account, as children are exposed to more language, they develop a stronger learned bias even when the "familiar" word is not previously known; Under the pragmatics account, as children are exposed to more language, they develop more skill in making social inferences, which would led to increased performance on the NN trials; And, under the bias and probabilistic accounts, maturational change could contribute to development in domain-general abilities, leading to a stronger disambiguation inference. Finally, the ability of children to succeed in the disambiguation tasks despite a range of imparements suggests that no single account likely describes a mechabnism that is both necessary and sufficient for the effect.

Experiment 1: ME and Vocabulary

The goal of Experiment 1 is to more directly explore the influence of vocabulary-related language experience on the disambiguation inference. Our meta-analysis points to a robust developmental increase in the strength of the disambiguation effect with age. While all four accounts are able to predict this change, only the overhypothesis account predicts that this increase should be related to vocabulary knowledge. In our meta-analytic analysis, we explored the relationship between vocabulary size and the magnitude of the disambiguation effect in the prior literature, but this analysis is limeted by the fact that vocabulary size is not measured for most studies in our sample. In Experiment 1, we therefore aimed to test the prediction that children with larger vocabularies should have a stronger disambiguation bias by measuring vocabulary size on a large sample of children who completed the disambiguation task. Consistent with the overhypothesis account, we find X.

Methods

Participants. A sample of 226 children were recruited at the Children's Discovery Museum of San Jose. 86 children were excluded because they did not satsify our planned inclusion criteria: within the age range of 24-48 months ($n = 13$), completed all trials ($n = 66$), exposed to English greater than 75% of the time ($n = 37$), and correctly answered at least half of the familiar noun control trials ($n = 55$). Our final sample included 140 children ($N_{\text{females}} = 87$.

Stimuli. The disambiguation task included color pictures of 14 novel objects (e.g., a pair of tongs) and 24 familiar objects (e.g. a cookie; see SI). Items in the vocabulary assessment were a fixed set of 20 developmentally appropriate words from the Pearson Peabody Vocabulary Test (**???**). These words were taken from 9 different domains, including professions, food, outside things, instruments, animals, classroom, shapes, verbs, and household items.

Design and Procedure. Sessions took place individually in a small testing room away from the museum floor. The experimenter first introduced the child to "Mr. Fox," a cartoon character who wanted to play a guessing game. The experimenter explained that Mr. Fox would tell them the name of the object they had to find, so they had to listen carefully. Children then completed a series of 19 trials on an iPad, 3 practice trials followed by 16 experimental trials. In the practice trials, children were shown two familiar pictures (FF) on the iPad and asked to select one, given a label. If the participant chose incorrectly on a practice trial, the audio would correct them and allow the participant to choose again. The child then progressed to the experimental trials. Half of the experimental trials were disambiguation trials and the other half were control trials presented in random order. Half of the experimental trials were Novel-Familiar (NF) noun pairs and the other half were Novel-Novel (NN). [How were N words introduced for NN trials?] Control trials differed from disambiguation trials in terms of X. Reaction times were measured from the onset of the target word. Children were only alloweed make one selection.

After the disambiguation task, children's vocabulary was measured in a simple vocabulary assessment. In the assessment, children were presented with four randomly selected images, and prompted to choose a picture given a label. Children completed 2 practice trials followed by 20 test trials. As in the disambiguation task, reaction times were measured from the onset of the target word, and children could only make one selection.

Data analysis.

Results and Discussion

Could be specific strength of particular word in the NF pairing

but we also get it for NN trials alone

Experiment 2: ME and Familiarity

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants.

age_group	mean_age	n
2	30.98684	38
3	40.98571	35
4	52.16216	37

We planned a total sample of 108 children, 12 per between-subjects labeling condition, and 36 total in each one-year age gorup. Our final sample was 110 children, ages Inf – -Inf months, recruited from the floor of the Boston Children's Museum. Children were randomly assigned to the one-label, two-label, or three label condition, with the total number of children in each age group and condition ranging between 10 and 13.

Materials. Materials were the set of novel objects used in de Marchena et al. (2011), consisting of unusual household items (e.g., a yellow plastic drain catcher) or other small, lab-constructed stimuli (e.g., a plastic lid glued to a popsicle stick). Items were distinct in color and shape.

Procedure. Each child completed four trials. Each trial consisted of a training and a test phase in a "novel-novel" disambiguation task (**???**). In the training phase, the experimenter presented the child with a novel object, and explicitly labeled the object with a novel label 1, 2, or 3 times ("Look at the *dax*"), and contrasted it with a second novel object ("And this one is cool too") to ensure equal familiarity. In the test phase, the child was asked to point to the object referred to by a second novel label ("Can you show me the *zot*?"). Number of labels used in the training phase was manipulated between subjects. There were eight different novel words and objects. Object presentation side, object, and word were counterbalanced across children.

Data analysis. We followed the same analytic approach as we registered in Experiment 1, though data were collected chronologically earlier for Experiment 2. Responses were coded as correct if participants selected the novel object at test. A small number of trials were coded as having parent or sibling interference, experimenter error, or a child who recognized the target object, chose both objects, or did not make a choice. These trials were excluded from further analyses; all trials were removed for two children for whom there was parent or sibling interference on every trial. The analysis we report here is consistent with that used in (**???**), though there are some slight numerical differences due to reclassification of exclusions.

err_type	n	pct
changed mind	2	0.0045455
exp err	2	0.0045455
interference	11	0.0250000
no choice	8	0.0181818
recog obj	4	0.0090909

Results and Discussion

As predicted, children showed a stronger disambiguation effect as the number of training labels increased, and as noise decreased with age.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.3076191	0.1046804	2.938650	0.0032965
age_mo_c	0.0464060	0.0112418	4.127972	0.0000366
times_labeled_c	0.4832010	0.1287155	3.754022	0.0001740
age_mo_c:times_labeled_c	0.0214303	0.0135810	1.577960	0.1145749

We analyzed the results using a logistic mixed model to predict correct responses with age, number of labels, and their interaction as fixed effects, and participant as a random effect. We centered both age and number of labels for interpretability of coefficients. Model results are shown in Table XYZ. There was a significant effect of age such that older children showed a stronger disambiguation bias and a significant effect of number of labels, such that more training labels led to stronger disambiguation, but the interaction between age and number of labels was not significant.

ME in Models of Word Learning

Basic statistical biases ("explaining away")

Regier (2005) model shows ME emergent

as noted by M. C. Frank, Goodman, Lai, and Tenenbaum (2009), Yu and Ballard (2007) model (IBM machine translation model #1, (**???**) for that; subsequently adapted by Nematzadeh, Fazly, and Stevenson (2012)) shows ME as well.

this is because any conditional probability model will show the same effect

In other words, Markman and Wachtel (1988)'s sense of a basic inductive bias will likely be present in a wide variety of different learning models.

What is the experience-dependence of ME in these models? In the M. C. Frank et al. (2009) model, the strength of the ME response scales with the strength of the familiar word's mapping; the same thing is true for the other models presumably.

Open question whether the actual differerence in a 2-year-olds' and a 4-year-olds' strength of representation of "ball" is what matters here?

M. C. Frank et al. (2009) model shows ME, in fact stronger than basic conditional probability. This is in part due to the use of the intention variable.

As a side note, the (**???**) no retention finding is shown in an even more pragmatic model: Smith, Goodman, and Frank (2013) model shows ME with no retention (though explanation in that model is a little implausible "because the speaker might not be committed to that label and is just using it as a matter of convenience.")

Primary point: No support here for overhypothesis building, which is suggested by 1) the bilingualism results. In order to fit the bilingual data, in general we'd have to assume that strength of individual representations in monolinguals and bilinguals was a driver, and this seems unlikely. 2) no support for E1 vocab findings unless the entire developmental trend is due to strength of the familiar word representations. In general, the strong — likely false — claim from all of these models is that the individual representation of the familiar object strength is the only locus for developmental/population-related change.

McMurray, Horst, and Samuelson (2012) model has ME emerge from the competition dynamics of a neural network.

> Thus, the selection of the novel object is dependent on the learning rule, but not because the network needs to learn something about that object/word. Rather, the weights between the known word/objects and the unused lexical units must decay, and the weights between the novel ones must not in order to create a platform upon which real-time competition dynamics can select the right object. A different type of weight decay (for example, if all weights decayed on each epoch) would not preserve the right form of the weight matrix. However, learning is not the whole story: this pattern of connectivity could not be harnessed in situation time without the gradual settling process represented by the inhibition and feedback dynamics. Moreover, the model's ability to learn from M.E. referent selection may also depend on this competition/feedback cycle. The model must select a single lexical unit and selectively amplify the novel object in order to

eventually turn a word-referent link created during M.E. referent selection into a known word by associating the novel object with the novel word over many instances. Thus, while as a real-time process mutual exclusivity is likely to impact learning, it is really more the product of learning than a mechanism of it.

This proposal is complicated but might capture the global and local dynamics in Experiment 1 & 2 better than others.

(**???**) deal with bilingual data by adding a direct ME-related penalty, not letting it be emergent.

General Discussion

References

Bion, R., Borovsky, A., & Fernald, A. (2012). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*.

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184–205.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition. Hillsdale, NJ: Erlbaum*.

de Marchena, A., Eigsti, I., Worek, A., Ono, K., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*(1), 96–113.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-bates communicative development inventories*. Paul H. Brookes Publishing Company.

Frank, M. C., Goodman, N. D., Lai, P., & Tenenbaum, J. B. (2009). Informative communication in word production and word learning. In *Proceedings of the 31st annual meeting of the cognitive science society*.

Golinkoff, R., Mervis, C., Hirsh-Pasek, K., & others. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–125.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markman, E., Wasow, J., & Hansen, M. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*(6619), 813–815.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831.

Mervis, C., Golinkoff, R., & Bertrand, J. (1994). Two-year-olds readily learn multiple labels for the same basic-level category. *Child Development*, *65*(4), 1163–1177.

Nematzadeh, A., Fazly, A., & Stevenson, S. (2012). A computational model of memory, attention, and word learning. In *Proceedings of the 3rd workshop on cognitive modeling and computational linguistics* (pp. 80–89). Association for Computational Linguistics.

Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology*, *16*(3), 337–348.

Preissler, M., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13–B23.

Quine, W. (1960). *Word and object* (Vol. 4). The MIT Press.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*(6), 819–865.

Smith, N. J., Goodman, N., & Frank, M. (2013). Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in neural information* *processing systems* (pp. 3039–3047).

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165.

model	n	term	estimate	z	p
Overall estimate	157	intercept	-0.18 [-0.47, 0.11]	-1.21	0.23
		age	0.03 [0.03, 0.04]	11.32	<.01
Typically-Developing populations (FN trials)	117	intercept	-0.33 [-0.71, 0.05]	-1.73	0.08
		age	0.05 [0.04, 0.05]	11.85	<.01
Typically-Developing populations (NN trials)	18	intercept	0.06 [-0.8, 0.93]	0.15	0.88
		age	0.03 [0.01, 0.04]	3.55	<.01
Multilingual populations (FN/NN)	12	intercept	0.05 [-0.78, 0.87]	0.11	0.91
		age	0.02 [0, 0.03]	1.77	0.08
Non-Typically-Developing populations (FN/NN)	10	intercept	-0.58 [-2.08, 0.92]	-0.75	0.45
		age	0.04 [0.01, 0.06]	3.15	<.01

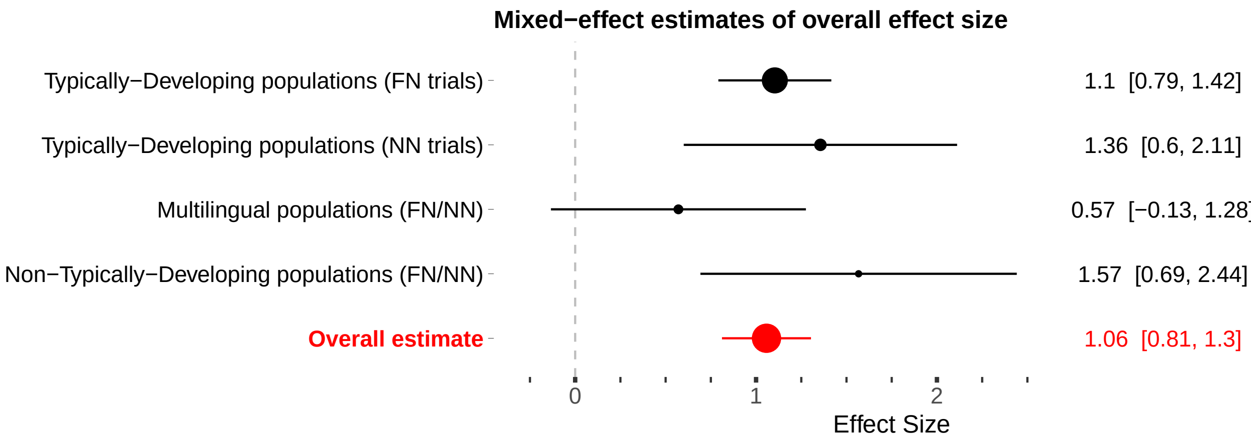

Figure 1. Mixed-effect effect size estimates for all conditions (red) and each of the four theoretically-relevant conditions in our sample. Ranges are 95% confidence intervals. Point size corresponds to sample size. FN = Familiar-Novel trials; NN = Novel-Novel trials.

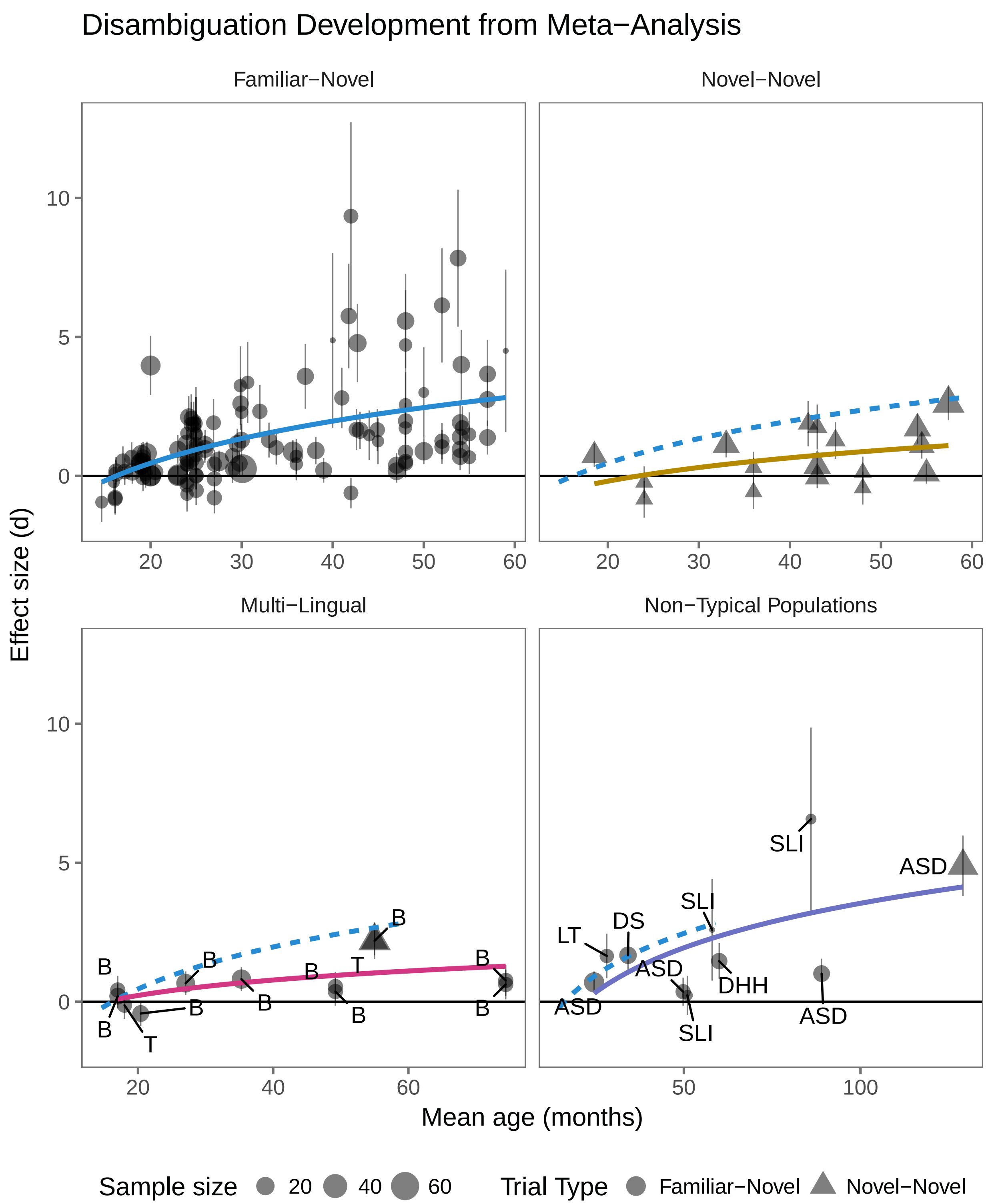

Figure 2. Developmental plots for each moderator. Ranges correspond to 95% confidence intervals. Model fits are log-linear. Note that the x-axis scale varies by facet.

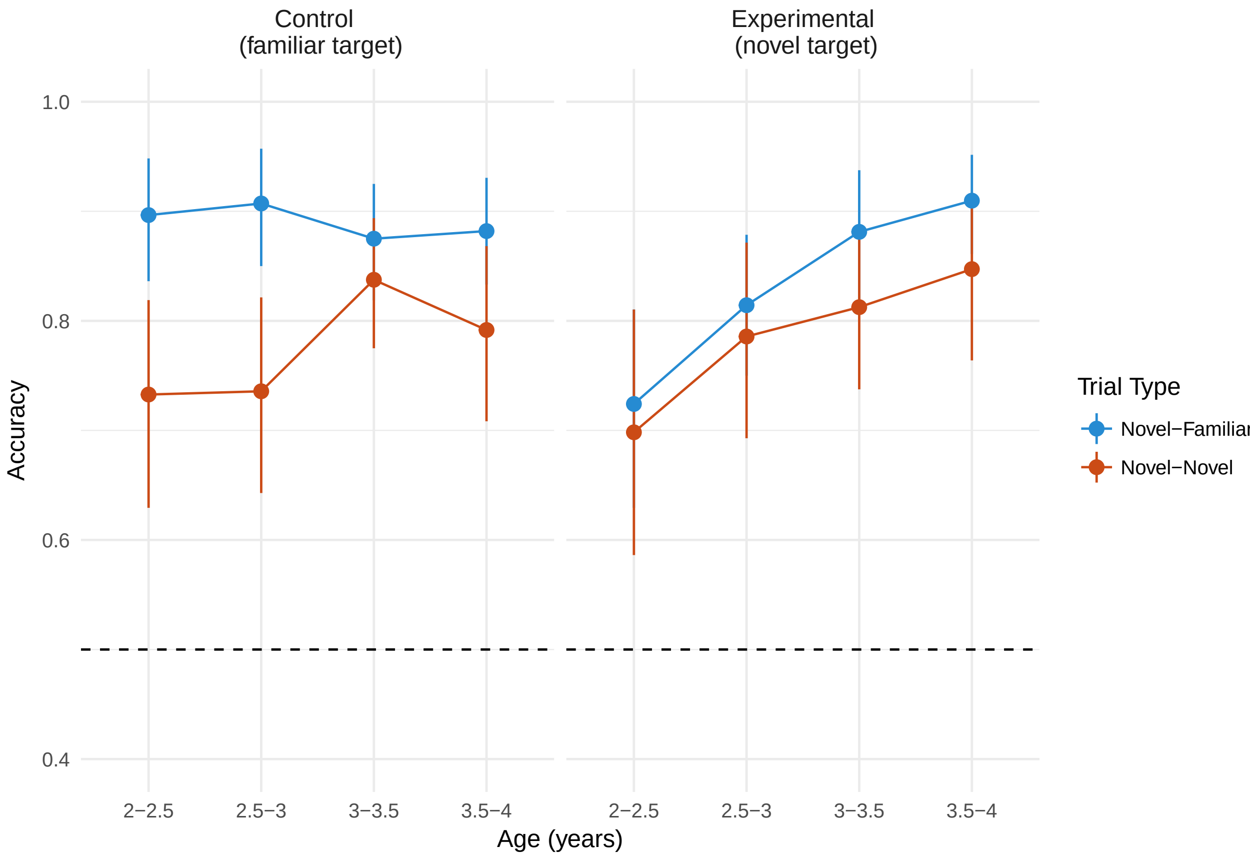

Figure 3. (#fig:accuracy_plot)Accuracy data for four trial types by half-year age bins. Blue corresponds to trials with the canonical novel-familiar paradigm, and red corresponds to trials with two novel alternatives, where the name of one is unambiguously introduced on a previous trial.

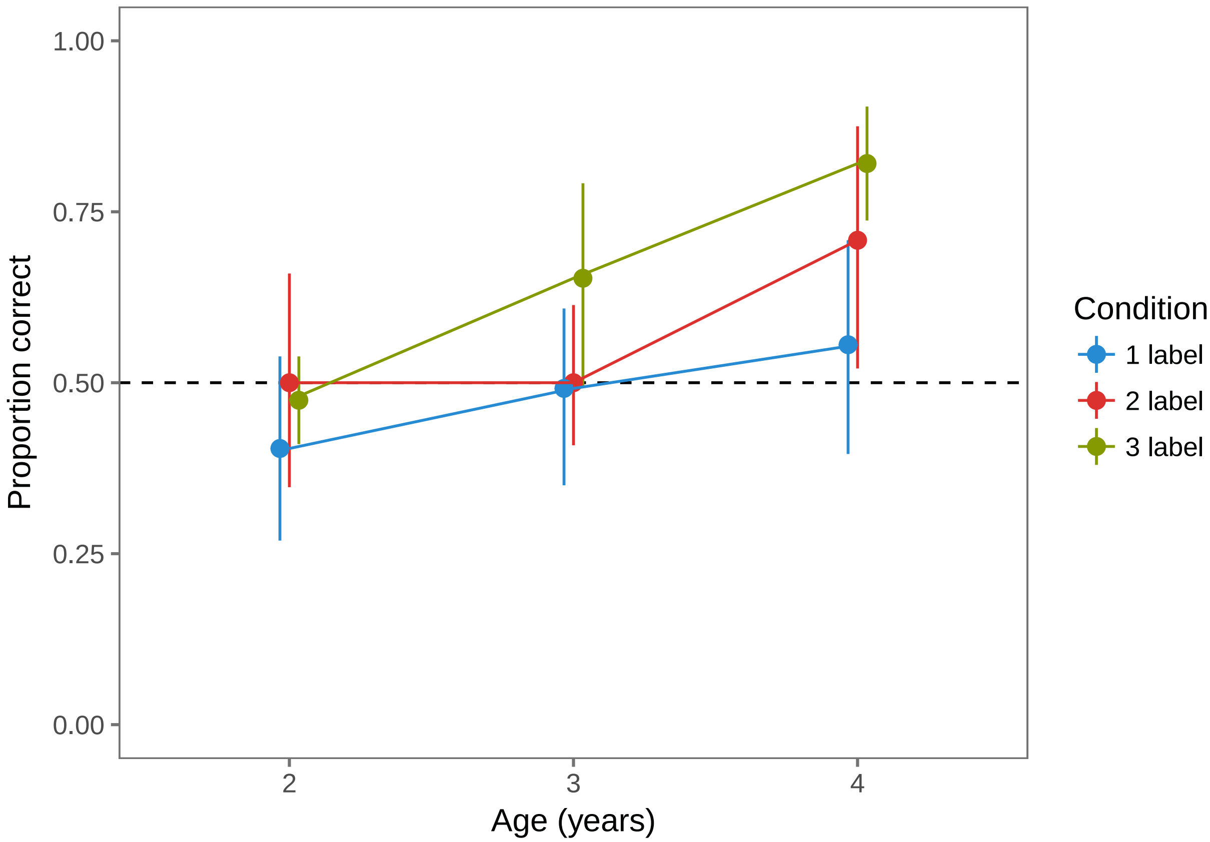
Proportion correct
1.00
0.75
0.50
0.25
0.00
2
3
4
Age (years)
Condition
1 label
2 label
3 label

Figure 4