How well do models of cross-situational word learning account for the learning of ambiguous words?

Sophie Regan[1]**, Isabelle Dautriche**[2]**, Mahesh Srinivasan**[1]
[1]Department of Psychology, University of California, Berkeley
[2]Centre de Recherche en Psychologie et Neurosciences, CNRS, Aix-Marseille Université
sophie_regan@berkeley.edu, isabelle.dautriche@cnrs.fr, srinivasan@berkeley.edu

Abstract

Existing theories of word learning largely focus on a learner's ability to learn a single meaning for a word despite the fact that many words have multiple meanings. Several computational models of cross-situational word learning have been proposed to explain how words are learned, but it is unknown to what extent they can learn ambiguous words with multiple meanings. Here, we present an experiment showing that adult learners are able to learn multiple meanings of novel ambiguous words in a cross-situational word learning paradigm, and are especially good at doing so when the meanings of the words are related (polysemous) rather than unrelated (homophonous). We evaluated the ability of ten different computational models of cross-situational word learning to explain the empirical data, and none were able to learn the ambiguous words as successfully as the adult learners. Moreover, because these computational models do not represent any semantic information, they are in principle unable to replicate the key difference between polysemous and homophonous word learning found in the study.

Keywords: Language Acquisition; Cross-Situational Word Learning; Lexical Ambiguity; Computational Modeling

Introduction

When a child hears a word in a given situation, there is ambiguity both regarding what the word might refer to (e.g., upon hearing "dog" in the presence of a dog and a leash, it is unclear which of these objects is being described), as well as how far the meaning should be generalized (e.g., if the word "dog" is meant to apply to a dog, could it also be used to describe a cat?). As has long been noted, these problems pose inductive challenges, requiring children to go beyond the evidence they have observed (Quine, 1960). Prior work suggests that children and adults overcome these challenges in part by learning the meanings of words *across individually ambiguous situations* (Smith & Yu, 2008). For instance, after hearing "dog" when a dog and a leash are present, a learner may be unsure about the meaning of the word, yet upon hearing "dog" another time in the presence of a dog and a bone, the learner may guess that the word refers to dogs because this was the only consistent referent across the word's usages.

While there is consensus that information about a word's meaning can be extracted from the environmental statistics of its use, theories differ regarding the mechanisms that support this process (Yurovsky & Frank, 2015). As a result of this divergence, there has been great interest in building computational models that explicitly implement distinct hypotheses regarding how words are learned across situations. For instance, according to associative models, learners update a matrix of possible word-referent associations each time they hear a word, strengthening associations with plausible referents and weakening associations with absent referents (Fazly et al., 2010; Kachergis et al., 2012). In contrast, in hypothesis-testing models, learners maintain only a single hypothesis about a word's meaning; each time they hear a word, this hypothesis is either verified or falsified (in the latter case, learners adopt a new hypothesis) (Medina et al., 2011; Trueswell et al., 2013). Others combine these ideas in hybrid models (Stevens et al., 2017).

A key–but largely overlooked–consideration in evaluating the ability of these computational models to explain human word learning is how well they account for the learning of ambiguous words–words that have multiple meanings (Yurovsky & Yu, 2008). While often treated as an edge case in language learning, most of the words that people learn are ambiguous. This includes cases of homophony (in which words have unrelated meanings, e.g., baseball vs. animal "bat"; estimated to apply to 7% of English words) and the far more prevalent cases of polysemy (in which words label multiple related senses of meaning, e.g., "chicken" animal vs. meat; estimated to apply to 84% of English words; (Rodd et al., 2002)). Although it has been claimed that some computational cross-situational models can learn ambiguous words (Fazly et al., 2010; Soh & Yang, 2021; Stevens et al., 2017), to our knowledge, no prior studies have directly compared model predictions in these cases against data from humans. This is a notable omission given evidence that even toddlers can learn multiple meanings for ambiguous words (Dautriche et al., 2018; Floyd et al., 2020; Rabagliati et al., 2010; Srinivasan & Snedeker, 2014; Srinivasan et al., 2017). Indeed, by some accounts, polysemous words in particular may be easier for children to learn than both homophones and unambiguous words, because they allow children to leverage their knowledge of one meaning of a word to make inferences about a word's other, related meanings (Floyd & Goldberg, 2021; Srinivasan & Rabagliati, 2021; Srinivasan et al., 2019).

In the current study, we present a pre-registered experimental study of adult learners' ability to learn multiple meanings of novel ambiguous words in a cross-situational learning paradigm in which each presentation of a word is referentially ambiguous. We show that adults are able to learn multiple meanings for a single word under these challenging conditions, and, additionally, that they learn the meanings

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Table 1: Word Stimuli.

Type	Meaning 1	Meaning 2
Homophonous	Broom	Ballet
Homophonous	Monkey	Bow
Homophonous	Pasta	Paws
Homophonous	Salt	Saddle
Homophonous	Worm	Glass (cup)
Polysemous	Crown	Wreath
Polysemous	Glass (material)	Mirror
Polysemous	Glasses	Goggles
Polysemous	Paper	Leaf
Polysemous	Pen	Feather

of polysemous words more easily than the meanings of homophonous words, mirroring prior findings of a "polysemy advantage" in children's word learning. Next, we present a systematic evaluation of the ability of ten computational models of cross-situational learning to account for the human data and find that the models present at best a limited ability to learn multiple meanings for each word. Moreover, because they do not include semantic information, these models are by definition unable to account for observed differences between the learning of polysemous and homophonous words.

Methods

Methods and analysis were preregistered at *https://aspredicted.org/2jrt-d338.pdf*.

Participants

A preregistered sample of 140 adult participants was recruited via Prolific and completed the study online. All participants were monolingual English speakers residing in the US. Participants were excluded if more than 50% of their responses were excluded ($N = 0$) or if they failed any of the catch trials ($N = 13$). This was a less stringent exclusion criterion than the preregistered criteria, which was originally to remove all participants who answered too quickly or too slowly on any single trial (which led to an overly high exclusion rate ($N = 50$)).

Procedure

Participants were randomly assigned to one of two conditions: a polysemous condition or a homophonous condition. In the polysemous condition, participants were tasked with learning novel words that had two related meanings. In the homophonous condition, participants learned novel words with two unrelated meanings. Table 1 shows the words used for each condition. The meanings of polysemous (and homophonous) words corresponded to meanings that were polysemous (or respectively, homophonous) in another language (French or Spanish) but critically *not* in English. Words were categorized as polysemous if the meanings were perceived as semantically related and homophonous if not. Validating these designations, a separate group of 30 participants rated the word pairs for conceptual similarity. Polysemous meanings were rated as more conceptually related than homophonous meanings (polysemous *mean* = 49.29, *sd* = 19.75 vs. homophonous *mean* = 4.52, *sd* = 1.15 on a scale of 1 to 100).

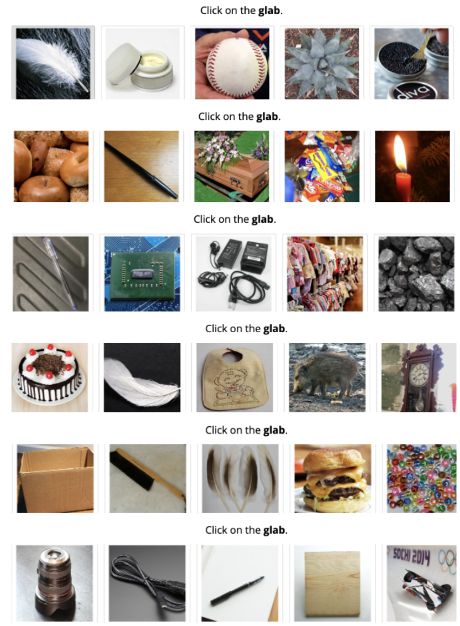

Figure 1: Example learning phase for a word meaning *Pen* and *Feather*.

In both conditions, participants also completed control trials in which they learned words with a single meaning. In the polysemous condition, word meanings on the control trials were taken from the list of homophonous words, with one of the meanings from each homophonous word being randomly selected to be the sole meaning of a control word. The reverse (forming control words from polysemous items) was true for the homophonous condition. This was done to ensure that one set of words was not *a priori* more difficult than the other.

In each condition, participants were told they would be learning new words in a pretend foreign language. Participants learned either the five homophonous words or the five polysemous words in a random order. Critical words (those with two meanings) alternated with control words (those with a single meaning). For each word, we measured learning both in participants' performance in a *learning phase* and in a *test phase*, as explained below.

Evidence about each word was presented on six consecutive learning trials in the learning phase. Each learning trial contained one target object (a referent of an actual meaning of the word) and four distractors. Participants were presented with a nonce word (e.g., *dax*) and told to select the word from the images. For critical words, each word had two meanings, and each meaning was presented on three of the six learning trials, in a pseudo-random order. For control words, each

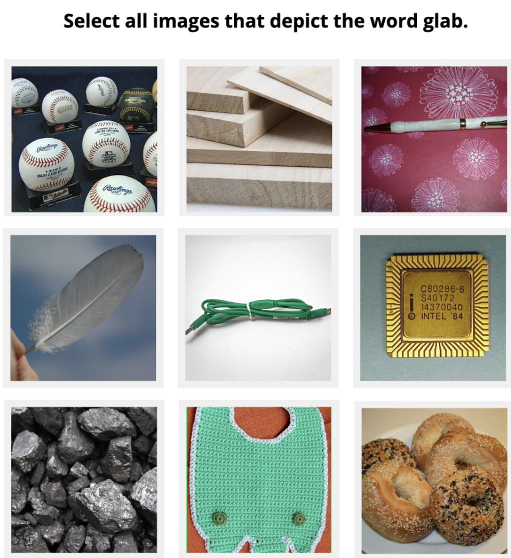

Figure 2: Example test trial for a word meaning *Pen* and *Feather*.

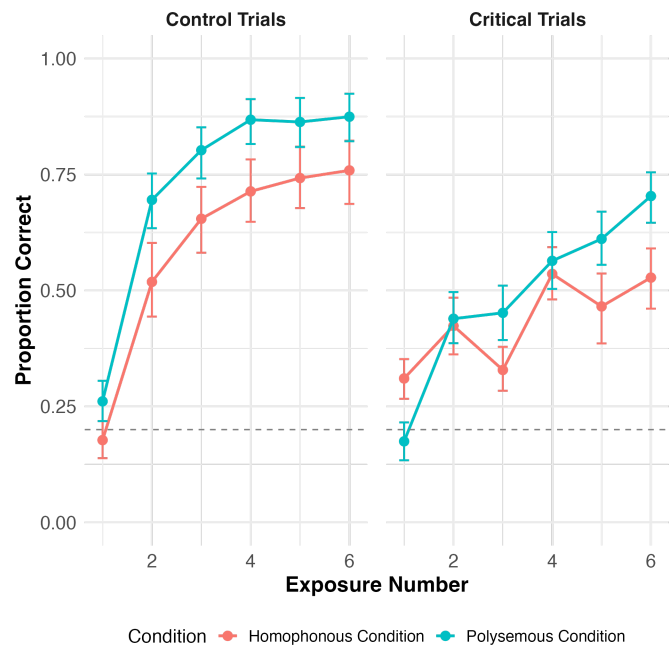

Figure 3: Results for the Learning Phase. Error bars represent 95% confidence intervals. The dashed line indicates chance performance.

word had a single meaning, whose referent was presented on each learning trial. Distractors were never repeated, and the target meanings were represented by different instance images on each exposure. Figure 1 shows a typical set of learning trials for a word meaning *pen* and *feather*.

For each critical word, after the learning phase, which consisted of the six learning trials, participants entered the test phase to probe whether they had successfully learned its multiple meanings. This was measured via a single test trial where participants saw images representing the critical word's two target meanings (one for each meaning) and seven of the distractor meanings that had been present during the learning phase. In the test phase for control words, participants instead saw the single target meaning and eight distractors. For both critical and control words, participants on test trials were told to select all of the images that depicted the target word. They were instructed that there would be at least one correct answer and that there could be more (See Figure 2 for a sample trial). Distractors and target meanings in the test trials were represented by new instance images that the participants had not previously seen. Participants also completed three randomly interspersed catch trials where they were explicitly told to select common objects (e.g., "Select all of the images that show a cat.").

Images were taken from the THINGS database (Hebart et al., 2019). For each concept, the seven most representative images were selected from the database (based on the database's norming data). On each trial, one of the images of the concept was selected randomly without replacement.

Analysis

Individual responses were excluded from analysis if the response time was shorter than 500 ms or longer than 30,000 ms ($N = 172$ responses out of $9,660$). Responses for entire words were removed if the participant was too fast or slow on 3 or more of the 6 learning trials ($N = 6$). As pre-registered, we hypothesized that accuracy would be higher on control words compared to critical words for both the learning phase and test phase because the critical words required participants to learn two meanings, rather than one. Further, based on prior work showing that learners show an advantage at learning polysemous words, we hypothesized that accuracy would be higher on critical words in the polysemous condition (compared to the homophonous condition). As we expected little difference in performance across conditions in the control trials, but we expected a difference in the test trials, we predicted a significant interaction effect of condition and trial type.

Participants' responses on both the learning phase and the test phase were analyzed with separate mixed-effects logistic regression models predicting accuracy as a function of condition and trial type (control vs. critical) with a random intercept for participant (Formula: accuracy $\sim$ condition $\times$ trial_type $+ (1|\text{participant})$). During the test phase, the response was coded as correct if all correct objects were selected (two in the critical trials and one in the control trials) and *no* incorrect objects were selected.

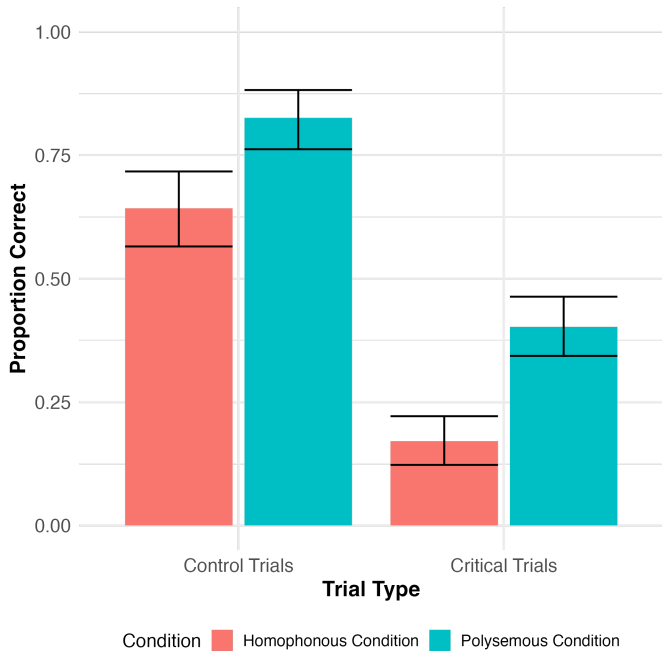

Figure 4: Results for the Test Phase. Error bars represent 95% confidence intervals.

Results

Results for the learning phase and the test phase are shown in Figures 3 and 4, respectively. On critical trials, participants learned two meanings for polysemous words in the polysemous condition or two meanings for homophonous words in the homophonous condition. On control trials, participants learned a single meaning for homophonous words in the polysemous condition or a single meaning for polysemous words in the homophonous condition. As hypothesized, accuracy on the critical trials was higher in the polysemous condition than in the homophonous condition in both the learning and the test phases. In the learning phase, participants were significantly more accurate on the control trials ($\beta = 1$, $p = .005$) and in the polysemous condition ($\beta = -3.2$, $p < .001$). Further, there was a significant interaction between trial type (control vs. test) and condition ($\beta = -.31$, $p < .005$). Interestingly, accuracy was also higher in the polysemous condition than in the homophonous condition on *control trials* as well. This means that the polysemous words were learned *more* easily than the homophonous words when the words had two meanings but were learned *less* easily when they had a single meaning.

Similarly, in the test phase, participants were significantly more accurate on the control trials ($\beta = .5$, $p < .005$) and in the polysemous condition ($\beta = -.92$, $p < .001$), but there was no significant interaction between condition and trial type ($\beta = .34$, $p = .31$). On control trials, participants chose the correct referent and no distractors on 82.59% [CI: 76.2%, 88.2%] of trials in the polysemous condition and on 64.27% [CI: 56.5%, 71.1%] of trials in the homophonous condition. On critical trials, participants chose both correct referents and no distractors on 40.24% [CI: 34.4%, 46.4%] of trials in the polysemous condition and on 17.07% [CI: 12.3%, 22.1%] of trials in the homophonous condition.

Discussion

As predicted, words with a single meaning were learned more readily than those with two meanings. Further, and also as hypothesized, among words with two meanings, polysemous words were learned more easily than homophonous words. Notably, despite the difficulty of the homophonous condition, participants were above chance on each learning trial. However, on the test trials when they had to select all meanings of the word, they were able to choose both correct meanings and no incorrect meanings less than a quarter of the time (17.07% of trials), suggesting that in many cases, they did not actually learn that the word had multiple meanings. In contrast, participants chose both meanings of the words (and no incorrect meanings) more than twice as often on critical trials in the polysemous condition (40.24% of trials), suggesting that many participants recognized that these words had multiple meanings. This success is notable given that participants experienced each meaning only three times and encountered significant ambiguity about the words' meaning(s). Overall, the results of this experiment align with previous research suggesting a polysemy advantage in word learning (Floyd & Goldberg, 2021; Srinivasan & Rabagliati, 2021; Srinivasan et al., 2019).

Interestingly, accuracy was significantly higher on control unambiguous words in the polysemous condition than in the homophonous condition even though a previous norming study had determined that these words were equally easy to learn. Notably, the critical meanings in the polysemous condition were the unambiguous meanings in the homophonous condition, and vice versa, so this effect is not driven by one set of meanings being a priori easier to learn. We speculate that learning words with two related meanings (in the polysemous condition) may have drawn the participants' attention to relations between referents that made them better on unambiguous words. It is also possible that the increased difficulty of learning words with two unrelated meanings (in the homophonous condition) may have made participants worse at the more straightforward task of learning words with a single meaning.

An alternative explanation is that participants learned a single superordinate category (e.g., *glasses* and *goggles* as *eyewear*) rather than two distinct meanings. While this is possible, it seems unlikely, as some polysemous pairs (e.g., *pen* and *feather*) are not easily grouped under a common label.

Next, we investigate how well ten proposed models of cross-situational learning can account for the human data. Given that these models do not consider any semantic information, they are, by definition, unable to provide different predictions regarding the learning of polysemous and homophonous words. But we can still probe their ability to learn multiple meanings for a single word.

Comparison to Computational Models

To assess the performance of current models of cross-situational word learning, we simulated the performance of ten common models in the three domains we tested with human participants: learning words with a single meaning (control trials), two related meanings (polysemous trials), and two unrelated meanings (homophonous trials).[1] All models were adapted from those presented in Kachergis and Frank, 2021, and we use the best fit parameters from that paper. (All code can be found at *anonymised link*). Each model is introduced below, and in-depth summaries of the models are available in Kachergis and Frank, 2021.

Baseline Co-Occurrence: Simply tracks how often words and objects appear together during trials.

Familiarity- and Uncertainty-Biased (FU): Assumes that learners form associations between all presented words and objects, prioritizing pairs with prior co-occurrence (familiarity) and those with weak or diffuse associations (uncertainty) (Kachergis et al., 2012).

Strength-, Uncertainty-, and Novelty-Biased: Versions of the FU model. The strength-biased model prioritizes familiar associations by excluding an uncertainty bias. The uncertainty-biased model lacks a strength bias. The novelty-biased model replaces entropy with a novelty measure (Kachergis et al., 2012).

Probabilistic Associative Model (PA): Represents the meaning of each word as a probability distribution over objects, updated incrementally across trials. Associations grow more quickly for familiar pairings unless another word strongly competes for the same object, reflecting a competitive learning process (Fazly et al., 2010).

Bayesian Decay (BD): Updates word-referent probabilities incrementally, reinforcing co-occurring pairs based on a likelihood function while penalizing non-co-occurring pairs (Kachergis & Frank, 2021).

Rescorla-Wagner (RW): Adapts associations between words and objects using prediction-error learning (Kachergis & Frank, 2021; Wagner & Rescorla, 1972).

Propose-but-Verify (PbV): Randomly selects a referent for any word without a remembered referent, then recalls and verifies the previously proposed referent, strengthening the association if verified and replacing it with a new hypothesis if not (Trueswell et al., 2013).

Pursuit: Adjusts associations dynamically by strengthening the strongest word-referent pair if present, weakening it if absent, and pairing novel words with weakly associated referents, using thresholded associations for referent selection (Stevens et al., 2017).

[1]Given that the models do not process any word-related information, the phrase "polysemous trials" in this section simply refers to the models' responses to the trials that were completed by the participants in the polysemous condition, and vice versa for "homophonous trials."

Table 2: Model Results RMSE.

Model	Control	Polysemous	Homophonous
Baseline	.25	.18	.12
FU	**.10**	.10	**.08**
Str	.12	.15	.11
Unc	.25	.15	.19
Nov	.14	.13	.10
PA	.21	.11	.14
BD	.25	**.08**	**.08**
RW	.28	.16	.11
PbV	.46	.35	.25
Pur	.18	.23	.14

Figure 5 shows average model performance on the learning trials, and table 2 shows the root mean squared error (RMSE) of each model vs. human performance on control, polysemous, and homophonous trials for the learning phase. As expected, no models were able to match the differences in polysemous and homophonous words that humans displayed since they are in principle unable to distinguish between the learnability of individual words at all. However, most of the models appeared to successfully learn two mappings at the same time (with the exception of Propose but Verify). Different models most closely approximated human performance in the three domains we tested. The FU model performed most similarly to human participants on the control trials ($RMSE = .1$) while the Bayesian Decay model performed most similarly to the human performance on the polysemous condition ($RMSE = .08$), and the Bayesian Decay and FU models performed most similarly to human performance on the homophonous condition ($RMSE = .08$).

Critically, participants completed not only a learning phase where they selected one object at a time but also a more stringent test phase, which probed their ability to choose all of the meanings of the word *and* no distractors. While modeling of the learning phase is straightforward since the model can presuppose that there is one correct answer, the test phase was more open-ended, as participants were told that there was at least one and potentially more than one correct answer. There are different ways to simulate model performance for the test phase. In our first approach, we took the association matrix produced by the simulation of the learning trials. Then, we simulated the test trials by selecting seven distractor objects at random (or eight in the control trials) and the two target objects (or one in the control trials). Objects were selected based on a softmax transformation of the association matrix values using a fitted temperature parameter, simulating probabilistic choice. Specifically, we row-normalized the association matrix and selected referents with probabilities proportional to their normalized association strengths. This yielded very poor performance: the most common outcome was for the models to select no objects at all on critical trials (between 40% and 70% of trials depending on the model), likely

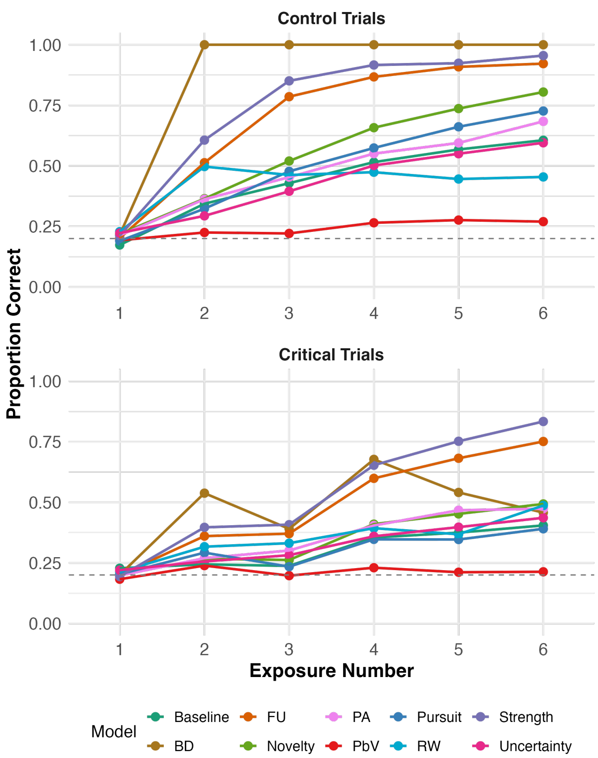

Figure 5: Model Results for the Learning Phase. The dashed line represents chance.

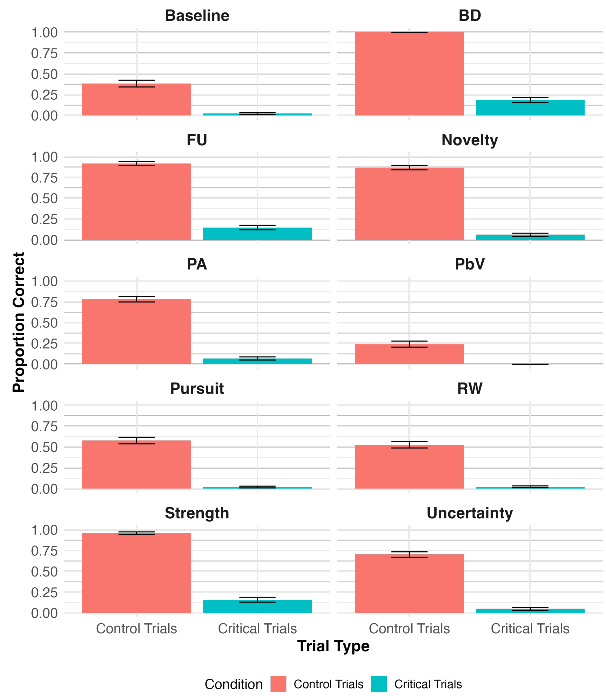

Figure 6: Model Results for the Test Phase.

because the magnitudes of values in the association matrices of the models were generally fairly low. The next most common outcome was selecting only one target meaning (between 10% and 40% depending on the model). In sum, on this approach to modeling the test phase data, the models virtually never demonstrated knowledge of the two meanings of the critical ambiguous words.

Next, we explored a secondary approach where we exponentiated the association matrix produced by the simulation of the learning trials by a parameter α (determined via differential evolution with the DEoptim package in R to minimize the sum of squared error between human and model performance) (Mullen et al., 2011). Finally, each referent was again selected as being a correct meaning with the probability of its value in the association matrix. Model performance on the test trial is shown in Figure 6. Despite relative success in the learning phase, none of the models we tested were able to match human performance on the test trial. On critical word trials, most of the models selected both correct referents (and no distractors) on less than 7% of trials despite selecting the correct referents during the learning phase at an above-chance rate. Notably, the Bayesian Decay, FU, and Strength models succeed at rates (24.1%, 14.6%, and 18.9%, respectively) close to human participants' performance on critical polysemous (43.2%) and homophonous word trials (16.9%). However, they also significantly overshoot human performance on the control trials ($>$ 99%, 91.8%, and 97.4%, respectively, as opposed to 85.2% for human polysemous condition control trials and 62.4% for human homophonous condition control trials), which indicates that they do not perform in a very human-like way on this particular task.

General Discussion

We have demonstrated a cross-situational word learning task where humans and computational models widely diverge. In our task, adult participants were able to learn multiple meanings of novel ambiguous words, particularly when those words were polysemous and labeled multiple related meanings. Moreover, a significant subgroup of participants (about 40%) demonstrated knowledge that the novel polysemous words had two meanings in the more stringent and open-ended test phase of our task. By contrast, although many of the computational models demonstrated success in the learning phase of our paradigm–where they could assume that only one of the presented images was correct–they were in general unable to match human performance in the test phase where they could not assume there was only one correct answer and instead had to select all correct meanings and no incorrect meanings. Moreover, although our participants were significantly better at learning the novel polysemous words than the novel homophones–in both the learning and test phases–the cross-situational learning models are in principle unable to account for these differences.

References

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, *12*(3), 231–251.

Dautriche, I., Fibla, L., Fievet, A.-C., & Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive psychology*, *104*, 83–105.

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

Floyd, S., & Goldberg, A. E. (2021). Children make use of relationships across meanings in word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 29.

Floyd, S., Goldberg, A. E., & Lew-Williams, C. (2020). Toddlers recognize multiple meanings of polysemous words. *Proceedings of the Annual Conference of the Cognitive Science Society*, 2799–2804.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS one*, *14*(10), e0223792.

Kachergis, G. (2012). *Mechanisms for cross-situational learning of word-referent mappings: Empirical and modeling evidence* [Doctoral dissertation, Indiana University].

Kachergis, G., & Frank, M. C. (2021). A large-scale comparison of cross-situational word learning models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic bulletin & review*, *19*, 317–324.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014–9019.

Mikolov, T. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, *3781*.

Mullen, K., Ardia, D., Gil, D. L., Windover, D., & Cline, J. (2011). Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, *40*(6), 1–26.

Quine, W. V. O. (1960). Word and object. new edition. *Cambridge, Massachusetts: The*.

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition*, *117*(1), 17–37.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language*, *46*(2), 245–266.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Soh, C., & Yang, C. (2021). Memory constraints on cross situational word learning. *Proceedings of the annual meeting of the cognitive science society*, *43*(43).

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11–24.

Srinivasan, M., Berner, C., & Rabagliati, H. (2019). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*, *148*(5), 926.

Srinivasan, M., & Rabagliati, H. (2021). The implications of polysemy for theories of word learning. *Child Development Perspectives*, *15*(3), 148–153.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive psychology*, *62*(4), 245–272.

Srinivasan, M., & Snedeker, J. (2014). Polysemy and the taxonomic constraint: Children's representation of words that label multiple kinds. *Language Learning and Development*, *10*(2), 97–128.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive science*, *41*, 638–676.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126–156.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in pavlovian conditioning: Application of a theory. *Inhibition and learning*, 301–336.

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

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. *Proceedings of the annual meeting of the cognitive science society*, *30*(30).