# Learning Absolute Meaning from Variable Exemplars

Crystal Lee (crystal.lee@rochester.edu)

and

Chigusa Kurumada (ckuruma2@ur.rochester.edu)

Department of Brain and Cognitive Sciences, University of Rochester

Corresponding author

Crystal Lee (crystal.lee@rochester.edu)

423 Meliora Hall, Department of Brain and Cognitive Sciences

University of Rochester, Rochester, NY, 14627

## Abstract

Linguistic theories often distinguish the semantic, or "literal", meaning of a word from the pragmatically enriched meaning communicated in context. This presents a logical problem to word learners who must acquire the semantic meaning through variable exemplars observed across contexts. This challenge for learners is particularly evident with regard to absolute gradable adjectives (e.g., *full, straight, wet*). Such adjectives are often used even when their literal (absolute) meaning does not apply (e.g., a 90% full cup can be "full" at a dinner table). Using a novel adjective, the current study shows that learners infer an absolute meaning by reasoning about how observed exemplars were generated given the relevant contextual constraints. Such inferential machinery allows learners to explain away visually represented variability and extrapolate an absolute meaning from the input consisting primarily of non-absolute exemplars. This highlights the importance of explaining away in word learning – specifically, in building semantic representations from contextually-conditioned word uses.


Keywords: word learning, absolute gradable adjective(s), inference, explaining away, pragmatics

## 1. Introduction

We use words to communicate. What we communicate, however, often goes beyond what words mean (Gadzar, 1979; Grice, 1975; Horn, 1972; Levinson, 2000). A now classic example of the English quantifier *some* (e.g., "Some students passed the test.") illustrates this. *Some* literally means *at least one*, but comprehenders often derive the interpretation "some *but not all* students" by leveraging their contextual knowledge and relevant alternative expressions (e.g., *all*). Contemporary theories of word meaning have observed the distinction of these two types of meanings, regarding the former type to be *semantic* (or literal) and the latter type to be *pragmatic*.

While this distinction is foundational and insightful, questions remain as to how these meanings are learned. Word learning is often considered to proceed primarily by associating words and observed referents (Clark, 1990; Gleitman, 1990; Heibeck et al., 1987; Kachergis, Yu & Shiffrin, 2012; Medina, Snedeker, Trueswell & Gleitman, 2011; K. Smith, Smith & Blythe, 2011; Pinker, 1994; Trueswell, Medina, Hafri & Gleitman, 2013; Vouloumanos, 2008; Yu & Ballard, 2007; Yu & Smith, 2007, 2011, 2012; Yu, Zhong & Fricker, 2012; Yurovsky, Smith & Yu, 2013, *inter alios*). But how do learners acquire the semantic meanings of words despite variability in how they are realized across contexts? While this question applies to most lexical items, including quantifiers and basic nouns, its depth can be better appreciated when we consider words that denote properties of objects. In this paper, we therefore examine learning of so-called absolute gradable adjectives.

The semantic meaning of absolute gradable adjectives (such as *full/empty, straight/bumpy, dry/wet*) is argued to have a maximum or minimum standard of comparison, independent of context, for a designated property (Rotstein & Winter, 2004; Kennedy & McNally, 2005; Kennedy, 2007). For instance, the semantic meaning of "full" can be roughly defined as "containing a maximal amount of content without spilling over". Exemplars of these adjectives observed by learners, however, often deviate from this meaning (e.g., a 90% full cup could be "full" when served at a dinner table), and vary across objects and contexts (e.g., a 70% full glass is very full when it contains wine but not when it contains water) (Kennedy & McNally, 2005). If learners are constructing hypotheses of the semantic meaning of absolute gradable adjectives based on these visually variable exemplars, they may fail to converge on the absolute meaning of the word. One key for learners to clue into the meaning is accompanying adverbs such as "completely" (Syrett & Lidz, 2010). These adverbs are not always available, however. And it still begs the question of how learners may come to acquire the semantic representations of a maximum or minimum standard of comparison.

The main goal of this paper is to propose and evaluate a new approach to the problem of learning absolute meanings based on contextually-conditioned and variegated exemplars. We hypothesize that learners infer the semantic meaning of absolute gradable adjectives through reasoning about how observed exemplars were generated given the relevant communicative constraints. For example, a learner maps the word "full" onto a 90% full cup at a dinner table together with the contextual understanding that the cup is maximally full for the current purpose *or else it will spill*. The same learner might also learn that a similar cup must be full to the brim when following a recipe. Learners can then causally attribute observed visual variability to latent contextual variables in order to explain them away, thus extrapolating an intended maximum standard across the exemplars. We call this the *explanation-based* account (Figure 1, left).

We contrast this account with a more classic associative approach, wherein an observed visual form-referent mapping is directly mapped onto a hypothesized meaning (Figure 1, right). For example, one proposal of cross-situational word learning suggests learners may store statistical distributions of all visual examples observed (Yu, 2008). Alternatively, they may rank their hypotheses and singularly pursue the most likely candidate ("propose but verify", Trueswell, Medina, Hafri & Gleitman, 2013). Irrespective of the details of the process, the most important predictor of word meaning under an associative model will be the observations of examples. For *full*, learners' initial hypothesis for the word meaning can be "maximally full" or "sufficiently full" depending on statistics in the observed exemplars. Learners must observe enough instances of visual examples that represent the maximum/minimum standard of comparison (e.g., a glass full to the brim) to learn the semantics of absolute gradable adjectives.

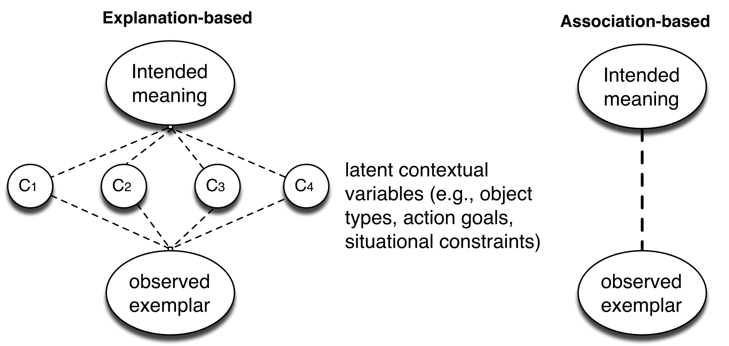


Figure 1. Schematic illustrations of association-based and explanation-based approaches.

In sum, the current study asks whether explaining away is active during word learning, as per the explanation-based hypothesis, providing a potential answer as to how literal meanings of absolute gradable adjectives are acquired though pragmatically-enriched exemplars. In two studies, we compare predictions of the explanation- and association-based approaches using a novel adjective whose meaning can include a maximum standard of comparison.

## 2. Experiment 1

We taught adult English speakers a novel gradable adjective, *pelty*, with absolute and non-absolute exemplars. This adjective is meant to describe a situation in which one object fits snugly into another, a naturalistic concept not lexically encoded in English (cf. in Korean: Choi et al., 1999; Norbury et al., 2008). The experiment consisted of an Exposure phase and a Test phase that assessed participants' learned meaning of the word.

### 2.1 Participants

79 participants were recruited from Amazon Mechanical Turk (https://www.mturk.com/mturk/). These participants were self-identified monolingual, English-speaking adults. Participants were randomly assigned to either With-Explanation or No-Explanation conditions. We excluded two participants due to poor catch trial performance and 14 participants due to past participations in similar experiments. All remaining 63 participants (30 in With-Explanation and 33 in the No-Explanation conditions) were included in our analyses.

### 2.2 Stimuli

Our Exposure stimuli consisted of videos illustrating uses of *pelty* paired with descriptions. The Exposure phase employed a total of 24 scenes, consisting of 6 unambiguously tight-fitting, 6 unambiguously loose-fitting, and 12 ambiguously tight-fitting scenes (Figure 2). The 12 ambiguously tight-fitting scenes were constructed from 6 videos, each repeated twice while paired with a different description. A total of 18 unique videos were created by crossing six object types (T-shirt, bracelet, bookshelf, shoe, laptop case, card) with the three degrees of tightness of fit: unambiguously tight-fitting, unambiguously loose-fitting, and ambiguously tight-fitting. Unambiguously tight- or loose-fitting exemplars were labeled as *pelty* and *not pelty*, respectively. For the two instances per object of a visually ambiguous video, one description labeled it as *pelty* while the other description labeled it as *not pelty.*

While the videos were identical across the two participant groups, the descriptions provided along with the videos were not. In the With-Explanation condition, the descriptions for visually ambiguous exemplars provided contextual explanations for the intended use of the word as a maximum standard absolute gradable adjective (e.g., "I want to wear this shoe with thick socks. This shoe is *pelty*."). Conversely, in the No-Explanation condition, the narration provided information about the object, but did not contain contextual explanations (e.g., "This is my favorite shoe. It's very comfy. This shoe is *pelty*".)

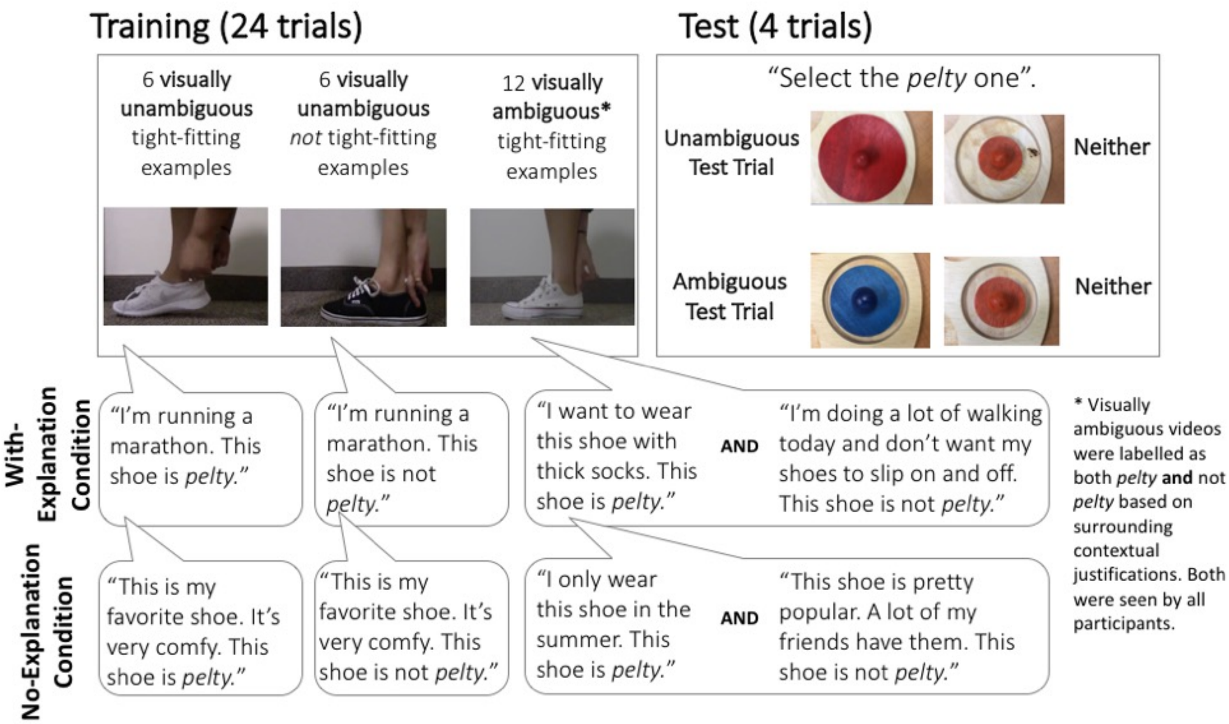


Figure 2. Experimental stimuli in the With-Explanation and the No-Explanation conditions of Experiment 1.

Our Test phase was modeled after Syrett et al. (2010), containing two kinds of novel objects (cylinders and puzzle pieces). There were two types of Test trials, in which participants were prompted to select an item in a three-alternative forced choice task. In the Unambiguous trials, these three choices were an unambiguously tight-fitting object (the tighter-fitting option), an unambiguously loose-fitting object (the less tight-fitting option), and "Neither" printed next to them. In the Ambiguous trial, the choices were a slightly (ambiguously) tight-fitting object (the tighter-fitting option), a largely loose-fitting object (i.e., more room around the edges, the less tight-fitting option), and "Neither" (top right of Figure 2). Participants saw a total of four trials (cylinders or puzzle pieces * Unambiguous or Ambiguous trials) presented in random order.

### 2.3 Procedure

Participants in both the With-Explanation and the No-Explanation conditions were instructed that a foreign-exchange student would be teaching them a new word, *pelty*, that is not used in American English. During the Exposure phase, participants watched 24 videos as described above in a randomized order and answered the question "Is this *pelty*?" at the end of each video by selecting a radio button (labeled as "yes" or "no"). This was to ensure that participants were receiving the contextual explanations presented via audio. In the Test phase, participants saw two still images and the printed word "Neither" and were prompted to "select the *pelty* one" (Figure 2).

### 2.4 Planned analysis of the data

First, we sought to determine if the participants had learned the meaning of *pelty* to include a maximum standard of comparison (such as *full*). If they had, then they should provide a greater response of 1) the tighter-fitting option in the Unambiguous trials; and 2) "Neither" in the Ambiguous trials because an object has to be maximally tight-fitting to be referred to as *pelty* when there is no relevant context present. If participants assigned a more comparative, as opposed to absolute, meaning (e.g., "tighter-fitting of the two"), they should provide a greater response of the tighter-fitting option in both the Unambiguous and Ambiguous trials.

Second, we sought to test if the between-subject context conditions had a significant effect on the learned semantic representations. Under the explanation-based account, the contextual explanations should allow participants to use their causal understanding of the context to explain away visual variability across absolute and non-absolute exemplars. In the absence of such contextual contributions (i.e., the novel objects seen in the Test phase), a maximum standard should apply. On the other hand, under the associative-based account, participants' learning of a maximum standard is primarily dependent on exposure to absolute exemplars (e.g., maximally tight-fitting objects). In that case, we do not expect to see any effect of the conditions because they did not differ in the number or content of absolute exemplars.

We planned to compare participants' response choices during Test using mixed-effects logistic regression (Breslow & Clayton, 1993; Jaeger, 2008). All analyses were conducted using the *glmer* function of the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) in R (R Core Team, 2016). We coded the responses as binary variables: "Neither" = 1 and all others = 0[1]. The analysis crossed condition (1 = With-Explanation vs. 0 = No-Explanation) with the type of Test trial (1 = Ambiguous vs. 0 = Unambiguous) as fixed effects and Test item kinds (cylinder vs. puzzle pieces) and subjects as random effects.

### 2.5 Results

Figure 3 shows the mean proportion of responses during the Test phase. In the Unambiguous trials, participants in both conditions chose the tighter fitting option to best denote *pelty*. Participants in the With-Explanation group did so more often (95.0%), compared to participants in No-Explanation group (69.7%). This suggests that participants in both With-

[1] As our outcome is a three-way categorical variable, a multinomial analysis would be more appropriate. However, repeated measures analysis of multinomial data with crossed random effects are computationally demanding and difficult to implement (see also Jaeger, Furth & Hilliard, 2012).

Explanation and No-Explanation conditions learned to associate the word *pelty* and the positive valency of fit-ness (for statistical analysis, see Supplemental Materials).

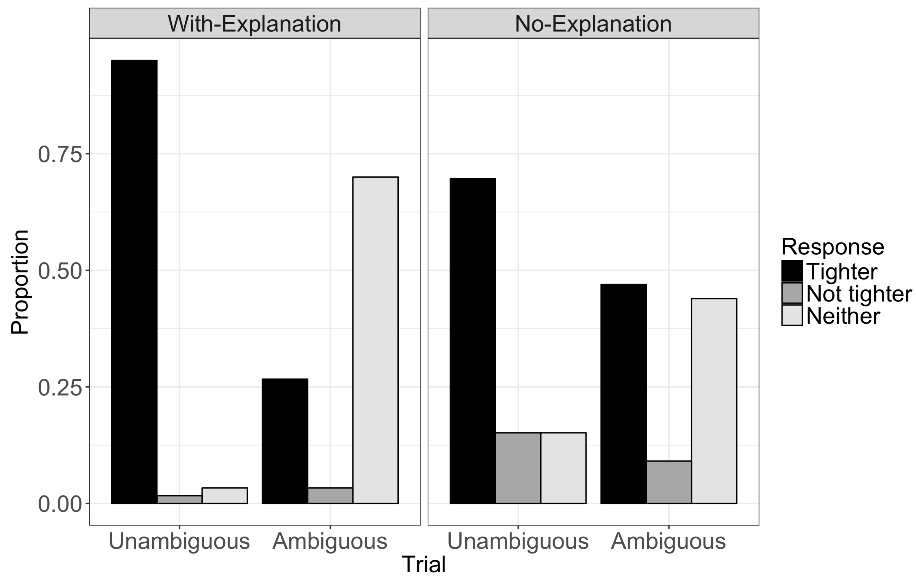


Figure 3. Proportion of responses collapsed across Test trials and participants in Experiment 1.

For the Ambiguous trials, consistent with the explanation-based account, participants in the With-Explanation condition were more likely to choose "Neither" (70.0%) compared to those who were in the No-Explanation condition (43.9%). Our analysis confirmed that there was a significant interaction of the proportion of "Neither" responses between trial (Unambiguous and Ambiguous) and condition ($\beta = 0.76$, $z = 3.23$, $p < .002$). Participants were more likely to choose "Neither" in the Ambiguous trial when provided contextual explanations during Exposure. Those in the No-Explanation condition were more willing to choose the more tight-fitting of the two.

Table 1. Model fixed and random effects predicting "Neither" responses in Experiment 1.

| **Experiment 1: Fixed Effects** | | | | |
|---|---|---|---|---|
| | **β** | **Std. Error** | **z-value** | **p-value** |
| Intercept | -1.28 | 0.27 | -4.61 | 4.01e-06 |
| With-Explanation Condition | -0.09 | 0.27 | -0.32 | 0.75 |

| | | | | |
|---|---|---|---|---|
| Ambiguous Trial | 1.66 | 0.26 | 6.3 | **2.89e-10** |
| With-Explanation Condition x Ambiguous Trial | 0.76 | 0.23 | 3.23 | **<.002** |
| | | | | |
| **Experiment 1: Random Effects** | | | | |
| **Groups** | | **Variance** | **Std. Dev.** | |
| Subject | (Intercept) | 1.09 | 1.05 | |
| Item | (Intercept) | 4.30e-10 | 2.07e-05 | |
| Observations: 252, Subjects: 63, Item Type: 2 | | | | |

Overall the results support the prediction of the explanation-based account. In Experiment 2, we put the explanation-based account to a stronger test and examined if participants could acquire the maximum standard even with a smaller number of absolute exemplars.

## 3. Experiment 2

Experiment 2 replicates Experiment 1 with a decreased number of absolute exemplars in the Exposure phase.

### 3.1 Participants

83 participants were recruited from Amazon Mechanical Turk. We used the same exclusion criteria as in Experiment 1. A total of 63 subjects were analyzed (30 in With-Explanation and 33 in No-Explanation condition).

### 3.2 Stimuli

The stimuli used were similar to those used in Experiment 1. However, the number of visually unambiguous items was halved (12 in Experiment 1, six in Experiment 2), and the number of visually ambiguous items increased by 50% (12 in Experiment 1, 18 in Experiment 2) (Table 2).

We created two lists to counterbalance the object types seen as ambiguous and unambiguous exemplars. For instance, some participants saw visually unambiguous exemplars of *book*, *shoe*, and *card* while others saw *laptop*, *shirt*, and *bracelet*. Six additional videos depicting visually ambiguous exemplars were created to complete the lists.

Table 2. Stimuli presentation differences across Experiments.

| | Visually unambiguous | | Visually ambiguous | | Total |
|---|---|---|---|---|---|
| | labeled as *pelty* | labeled as *not pelty* | labeled as *pelty* | labeled as *not pelty* | |
| Experiment 1 | 6 | 6 | 6 | 6 | 24 |
| Experiment 2 | 3 | 3 | 9 | 9 | 24 |

### 3.3 Procedure

The procedure was identical to Experiment 1.

### 3.4 Results

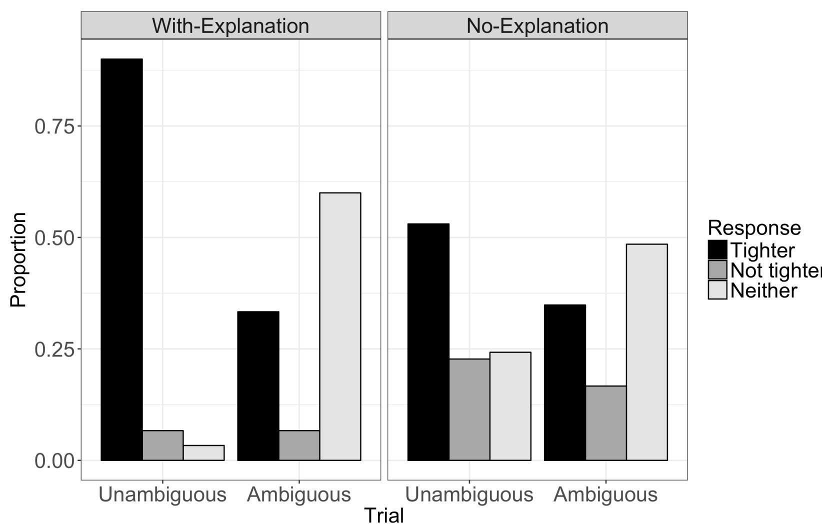


Figure 4. Proportion of responses collapsed across Test trials in Experiment 2.

Figure 4 illustrates the proportion of responses in the Test phase in Experiment 2. One notable difference between Experiments 1 and 2 is that participants in the No-Explanation condition were less likely to choose the tighter-fitting option in the Unambiguous trials (69.7% in Experiment 1 and 53.0% in Experiment 2). This may suggest that the increased number of the visually ambiguous exemplars made it difficult for them to reliably associate the form *pelty* and the meaning "tight-fitting".

We constructed a new mixed-effect regression model with data from both Experiments 1 and 2: Test trial type (Unambiguous and Ambiguous trials), Condition (With- and No-Explanation), and Experiment (1 and 2) were entered as fixed effects, and test item types and subjects were entered as random effects. The model revealed a significant interaction of the proportion of "Neither" responses between Test trial type (Unambiguous and Ambiguous) and condition (With- and No-Explanation) ($\beta = 0.76$, $z = 4.60$, $p < .0001$, Table 3). The three-way interaction including Experiments was not a significant predictor for "Neither" responses ($p < 1$), suggesting that participants in the With-Explanation condition in Experiment 2 were able to infer the meaning of *pelty* to be "maximally tight-fitting" as well as those in Experiment 1 despite the smaller proportions of prototypical exemplars seen (50% in Experiment 1, 25% in Experiment 2).

Table 3. Model fixed and random effects predicting "Neither" responses for both Experiments 1 and 2.

| **Experiment 1 and 2: Fixed Effects** | | | | |
|---|---|---|---|---|
| | **β** | **Std. Error** | **z-value** | **p-value** |
| Intercept | -1.24 | 0.19 | -6.37 | 1.86e-10 |
| With-Explanation Condition | -0.27 | 0.19 | -1.47 | 0.14 |
| Ambiguous Trial | 1.54 | 0.18 | 8.52 | **< 2.00e-16** |
| Experiment (1 or 2) | -0.04 | 0.19 | -0.24 | 0.81 |
| With-Explanation Condition x Ambiguous Trial | 0.76 | 0.16 | 4.60 | **4.15e-06** |
| With-Explanation Condition x Experiment | 0.19 | 0.19 | 1.01 | 0.31 |
| Ambiguous Trial x Experiment | 0.12 | 0.16 | 0.75 | 0.45 |
| With-Explanation Condition | 0.01 | 0.16 | 0.03 | 0.98 |

| | | | | |
|---|---|---|---|---|
| x Ambiguous Trial x Experiment | | | | |
| | | | | |
| **Experiment 1 and 2: Random Effects** | | | | |
| **Groups** | | **Variance** | **Std. Dev.** | |
| Subject | (Intercept) | 1.11 | 1.05 | |
| Item | (Intercept) | 5.09e-16 | 2.26e-08 | |
| Observations: 504, Subjects: 126, Item Type: 2 | | | | |

## 4. General Discussion

The apparent and considerable variability in word usage presents a significant challenge to theories that posit word learning as an associative process of mapping phonological forms of words onto observed exemplars. The current results clearly suggest that learners can explain away visually represented variability by taking into account contextual explanations (e.g., an imperfectly fitting shoe can be a referent of an adjective meaning "maximally tight-fitting" *when it is to be worn with a thick sock*). The effect of the contextual inference when faced with ambiguous exemplars was further magnified in Experiment 2, where learners received absolute exemplars only in 25% of the Exposure trials. Those who received contextual explanations still learned the meaning of *pelty* to include a maximum standard of comparison, while those who did not receive contextual explanations were significantly less certain about the semantics of the word.

The current approach is similar in spirit to theories of word learning as a process of actively inferring about a causal, generative process that produces the observed data (Frank, Goodman & Tenenbaum, 2009; Xu & Tenenbaum, 2007). Besides guiding attention to relevant aspects of context (Trueswell et al., 2016), speaker's action goals and contextual conditions of word use constitute part of the latent structure through which the observed data has been generated. Learners are tasked to model this process to infer a likely hypothesis (meaning) that generalizes to unseen data (Shafto, Goodman,  & Griffiths, 2014). Contextual explanations we provided in our experiment are considered to guide this inference. Mounting evidence from word learning (Carpenter, Akhtar & Tomasello, 1998; Tomasello & Barton, 1994) and social cognition (Csibra & Gergely, 2009; Meltzoff, 1995; Vouloumanos, Onishi, & Pogue, 2012) echoes this approach: word learners, including young children, are sensitive to the speaker's intentions conditioning her word use, which allows them to extrapolate intended form-meaning mappings even when not visually instantiated

It is important to note, however, that the explanation-based account does not preclude the necessity of observing absolute exemplars. In a follow-up experiment, all 24 videos during Exposure were visually ambiguous and associated with contextual explanations rendering them as *pelty* or *not pelty* (i.e., absolute exemplars were not seen in Exposure). In this experiment, participants did not infer a maximum standard of comparison of *pelty* (see Supplemental materials). This suggests that the process of explaining away must be supported by at least some representations of a maximum standard of comparison, drawing learners' attention to an intended endpoint of the relevant semantic scale. In our continuing work, we intend to delve further into the relationships between the associative and inferential machineries actively interacting in word learning.

The explanation-based account can be applicable to lexical classes beyond adjectives, and is potentially very powerful in explicating how children can infer a semantic meaning despite experiences of diverse pragmatic contexts. To exercise such inferences over an intended word meaning, children must understand various causal structures relevant to speaker intentions that generated a word in a given context. This is expected to require general cognitive maturation as well as exposure to the word in various contexts, potentially contributing to differential semantic representations between children and adults (e.g., Syrett, Kennedy, & Lidz, (2010) for absolute gradable adjectives). Future research should empirically assess the degree of variability in exemplars of a word meaning and how the resulting uncertainty can be reduced by taking contextual explanations into account.

## References


Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. DOI:10.18637/jss.v067.i01.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Child Language Development*, *15,* 17-29.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- to 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*, 315-330.

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language specific spatial categories in English and Korean. *Cognitive Development, 14*(2), 241-268. http://dx.doi.org/10.1016/

Clark, E. V. (1990). Speaker perspective in language acquisition. *Linguistics*, *28*(6), 1201 -1220.

Csibra, G., & Gergely,G. (2009). Natural Pedagogy. *Trends in cognitive sciences*, *13*. 148-53. 10.1016/j.tics.2009.01.005.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*(5), 578-585. https://doi.org/10.1111/j.1467-9280.2009.02335.x

Gadzar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. London: Academic Press.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

http://dx.doi.org/10.1207/s15327817la0101_2

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics.* New York, NY: Academic Press, 341–358.

Heibeck, T. & Markman, E. (1987). Word learning in children: An examination of fast-mapping. *Child Development, 58*(4), 1021-1034. DOI: 10.2307/1130543

Horn, L. (1972). On the semantic properties of the logical operators in English. Doctoral dissertation, UCLA, Los Angeles, CA.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1439-1449. http://dx.doi.org/10.1037/a0027862

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. DOI: 10.3758/s13423-011-0194-6

Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy, 30*(1), 1-45. DOI:10.1007/s10988-006-9008-0

Kennedy, C. & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language, 81*(2), 245-381.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT press.

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*, 9014–9019. DOI: 10.1073/pnas.1105040108

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. http://doi.org/10.1037/0012-1649.31.5.838

Norbury, H. M., Waxman, S. R., & Song, H. (2008). *Tight* and *loose* are not created equal: An asymmetry underlying the representation of *fit* in English- and Korean-speakers. *Cognition, 109*(3): 316-325. https://doi.org/10.1016/j.cognition.2008.07.019

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377 410. https://doi.org/10.1016/0024-3841(94)90347-6

R Core Team (2016). R: A Language and Environment for Statistical Computing [Computer software]. Retrieved from https://www.R-project.org/

Rotstein, C. & Winter, Y. (2014). Total adjectives vs. partial adjectives: scale structure and higher -order modifiers. *Natural Language Semantics, 12*(3), 259-288. https://doi.org/10.1023/B:NALS.0000034517.56898.9a

Shafto, P., Goodman, N., Griffiths, T. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71*, 55-89.

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

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*, 480–498. DOI: 10.1111/j.1551-6709.2010.01158.x

Syrett, K., Kennedy, C., & Lidz, J. (2010). The meaning and context in children’s understanding of gradable adjectives. *Journal of Semantics, 27*, 1-35. DOI:10.1093/jos/ffp011

Syrett, K. & Lidz, J. (2010). 30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives. *Language Learning and Development, 6*(4), 258-282. https://doi.org/10.1080/15475440903507905

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*(5), 639-650.

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

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., Gleitman, L. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition, 148,* 117-135. https://doi.org/10.1016/j.cognition.2015.11.002

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742. https://doi.org/10.1016/j.cognition.2007.08.007

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(32), 12933–12937. http://doi.org/10.1073/pnas.1121057109

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272. http://dx.doi.org/10.1037/0033-295X.114.2.245

Yu, C. & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*(13-15), 2149-2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420. http://dx.doi.org/10.1111%2Fj.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*, 165–180. DOI: 10.1111/j.1467-7687.2010.00958.x

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*, 21–39. DOI: 10.1037/a0026182

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross- situational statistical learning: Evidence from eye tracking. *Frontiers in Psychology*, *3*, 148. https://dx.doi.org/10.3389%2Ffpsyg.2012.00148

Yurovsky, D., Yu, C., Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science, 37*(5), 891-921. DOI: 10.1111/cogs.12035