Simultaneous Cross-situational Learning of Category and Object Names

Tarun Gangwani, George Kachergis, and Chen Yu
{tgangwan, gkacherg, chenyu}@indiana.edu
Department of Psychological and Brain Science, and Cognitive Science Program
1101 East 10th Street Bloomington, IN 47405 USA

Abstract

Previous research shows that people can acquire an impressive number of word-referent pairs after viewing a series of ambiguous trials by accumulating co-occurrence statistics (e.g., Yu & Smith, 2007). The present study extends the cross-situational word learning paradigm, which has primarily been used to investigate the acquisition of 1-to-1 word-referent mappings, and shows that humans can concurrently acquire both 1-to-1 and 1-to-many mappings (i.e., a category relation), even when the many referents of a single word have no unifying perceptual features. Thus, humans demonstrate an impressive ability to simultaneously apprehend hierarchical regularities in their environment.

Keywords: statistical learning; cross-situational learning; category learning; mutual exclusivity; language acquisition

Introduction

In order to make sense of their world, human infants must learn relationships between words and referents in their environment. Infants simultaneously come into contact with many diverse, novel objects and equally diverse words that name them. Thus, there is much potential for acquiring erroneous word-referent mappings, given only a single situation. Despite this, both infants and adults have a remarkable ability to learn many novel word-referent associations quickly and accurately. Cross-situational word learning (CSWL) studies give us insight into how people are capable of learning multiple word-referent associations from individually ambiguous situations. Previous CSWL studies have shown that both infants and adults are able to learn simple 1-word to 1-referent mappings with astonishing speed (Kachergis, Yu, & Shiffrin, 2009; Smith & Yu, 2008; Klein, Yu, & Shiffrin 2008; Yu & Smith, 2007). In adult studies, participants are typically instructed to learn which words go with which referents, and are then presented with a few consistently co-occurring objects and spoken pseudowords on each of a series of training trials. On every trial, each pseudoword corresponds to a particular on-screen object, but the intended referent is never indicated. In a typical cross-situational training block, participants attempt to learn 18 word-referent pairs from 27 twelve-second trials consisting of four spoken words and four displayed objects (*i.e.* a 4x4 design). On average, participants in this condition managed to learn half of the 18 pairs by relying on cross-situational statistics (Yu & Smith, 2007). Further studies have shown that human learning often reflects statistics manipulated during training such as pair frequency, contextual diversity (the diversity of other pairs each pair appears with over time), and within-trial ambiguity (the number of co-occurring words and referents per trial) (Kachergis, *et al*., 2009).

However, simple 1-to-1 mappings are only a subset of the types of word-referent relations that exist in natural languages. 1-to-many mappings include referents that have one common label shared among them, such as a category or concept label. For example, both an apple and a banana may be labeled 'fruit.' Learners must learn to map both the superordinate label ('fruit') and each basic level name ('banana' 'apple') to the appropriate referent. Even in a learning paradigm like the 4x4 cross-situational learning condition discussed above, which is simpler than the real world, it is difficult to imagine that learners consider all 16 possible pairings, as might be necessary to learn higher-order relations. Constraints such as mutual exclusivity (ME) can drastically reduce the complexity of such ambiguous situations by limiting the possible pairings to a single word for each object (and vice-versa). Consider Markman and Wachtel's study (1988), in which a child was placed in front of a learned object (ball) and an unlearned object (gyroscope) and was prompted to retrieve the 'toma.' While 'toma' could be another name for the ball, the child moves to the unlearned object, exhibiting ME. However, despite its power to speed learning, the strict use of ME as a constraint in cross-situational learning would also make it impossible to learn non-1-to-1 mappings.

To determine whether learners use the ME constraint when learning names for previously unknown objects, Yurovsky and Yu (2008) presented learners with ME-violating mappings in the CSWL paradigm. An ME-violating mapping is a word (or object) that is consistently paired with more than one object (or word). Participants were trained on 12 words and 18 referents, where 6 words were paired with 12 referents (*i.e.*, 2 referents per word), known as *double* words, and the other 6 words were paired 1-to-1 with the remaining 6 referents, known as *single* words. Participants had to decide how to manage two names that co-occur with one referent in the same set of trials. The results showed that participants had equal performance in learning both single and double words if each double word's two referents were interleaved rather than temporally separated (*i.e.* one referent was shown in the first half only

and the other appeared only in the last half). Moreover, learners acquired more than half of both early and late pairings; thus, some must have violated ME.

In contrast, Ichinco, Frank, and Saxe (2009) presented participants with a study to demonstrate ME as a guide to learning word-to-referent associations. Participants were shown an additional referent (or word, in a different experiment) on each trial, alongside four previously-seen word-referent pairs. Both groups received training on a standard cross-situational task, which was followed by further training. In this training, a new stimulus (word or object, between groups) was added on each trial alongside four pairs from the early training. Rather than forming a 1-to-2 mapping with the additional object (or 2-to-1, for the extra word) on each trial, participants learned 1-to-2 (or 2-to-1) relations on average for only one item and consistently favored mutually exclusive mappings.

Thus, depending on how ME-violating word-referent mappings are added to the cross situational paradigm, learners vary their use of the mutual exclusivity constraint. In the Yurovsky & Yu study, additional referents are presented in the absence of old ones: when participants hear a word and see two referents consistently co-occurring with it, they may be more likely to violate ME and form a 1-to-2 mapping. In Ichinco, *et al*.'s study, all 1-to-1 mappings from the early stage occur simultaneously with the new mappings. Participants may have failed to learn the new mappings due to blocking, a known associative learning effect in which a previously learned pairing interferes with the acquisition of a new pairing involving old stimuli.

In the present study, in order to eliminate biases that participants may adopt as a result of training order, we provide participants with cross-situational training that is simultaneously consistent with both 1-to-1 (basic-level name to referent) and 1-to-many (superordinate-level name to multiple referents) relationships on every training trial. For example, in Experiment 1, on each 3x2 trial, two words are basic-level names for the visible referents, and a third will act as a superordinate-level identifier. These superordinate level labels hence refer to four referents, including two that are not on present on a given trial. Thus, participants are simultaneously faced with two labels for each referent, and one of these labels also applies to three other referents. In Experiment 2, we give learners a more complex learning scenario: 4x2 trials on which two labels map 1-to-1 to the objects and each of the other two labels refer to one of the present objects, and three unseen objects. In both experiments, participants must learn the unique name for a referent as well as a label it shares with three other objects, some of which are not present on a given trial.

One block in each experiment is composed of objects that share some unifying perceptual feature like a hook or arrow shape, somewhat like objects belonging to natural categories. We test for generalization using stimuli in which the objects share each category's identifying feature from training, but the objects have different textures and shapes than those from training.

Experiment 1

In Experiment 1, participants were merely instructed to learn which words go with which objects—with no mention of the potential to form 1-to-many relations—and were then given a sequence of cross-situational training trials, each consisting of three words and two referents. Unbeknownst to learners, two of the words on each trial map 1-to-1 to one of the visible referents, and the third word refers to both objects, and also will consistently appear with two other referents during training. Participants must determine which words specify a 1-to-1 reference to an object and which word specifies a 1-to-many reference to both objects on each trial. If participants assume ME, participants will either learn 1-to-1 mappings or 1-to-many mappings, but not both.

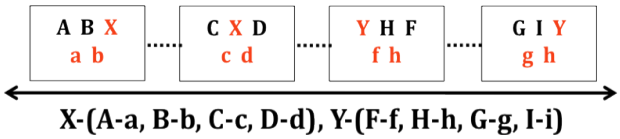

Figure 1: In Experiment 1, participants are trained on both 1-to-1 (e.g., A-a and B-b) and 1-to-many mappings (e.g., X-{a,b,c,d}) in the context of 3 words and 2 referents per trial. One word is the superordinate-level name that refers to both referents on each trial (shown in red).

In order to see if 1-to-many associations are facilitated by stimuli structure, subjects were trained on two different conditions (in three blocks in fixed order): **Block 1** was an *arbitrary category* condition, in which the objects had no obvious shared perceptual features but were consistently labeled by some other word. **Block 2** was a *natural category* condition, in which the objects in each category share a salient feature (e.g., a hook or arrow shape). **Block 3** was another *arbitrary category* condition (with different stimuli) to gauge attention shift after learning natural 1-to-many groupings. Given the salient features present in Block 2, performance in learning 1-to-many relationships will likely increase relative to Block 1, as participants' attention will be drawn to the 1-to-many relations due to the salient features acting as learning cues. Their performance on block 3 will indicate if this attentional shift is carried over from the natural category block.

Subjects

Participants were 33 undergraduates at Indiana University who received course credit for participating. None had participated in other cross-situational experiments.

Stimuli

Each training trial consisted of two objects shown on a computer screen and three pseudowords played sequentially. In each of the two arbitrary category conditions, the 12 referents were difficult-to-name, unrelated objects. For the

natural category condition, the 12 objects had one of three features protruding from the shape. The 45 computer-generated pseudowords are phonotactically-probable in English (e.g. "stigson"), and were spoken by a monotone, synthetic voice. 36 words are assigned to each referent, creating arbitrary word-object pairs which were randomly assigned to three sets of 12 1-to-1 mappings. One set of stimuli composed the natural category stimuli for the second block; the other sets composed of arbitrary strange objects for the first and third blocks. For the 1-to-many mappings, the remaining 9 pseudowords are assigned to three sets of four 1-to-1 mappings. Thus, in each block there are three groups (i.e., categories).

Words

Referents	X (12)				Y (12)				Z (12)			
	A	B	C	D	E	F	G	H	I	J	K	L
a	6	2	2	2	0	0	0	0	0	0	0	0
b	2	6	2	2	0	0	0	0	0	0	0	0
c	2	2	6	2	0	0	0	0	0	0	0	0
d	2	2	2	6	0	0	0	0	0	0	0	0
e	0	0	0	0	6	2	2	2	0	0	0	0
f	0	0	0	0	2	6	2	2	0	0	0	0
g	0	0	0	0	2	2	6	2	0	0	0	0
h	0	0	0	0	2	2	2	6	0	0	0	0
i	0	0	0	0	0	0	0	0	6	2	2	2
j	0	0	0	0	0	0	0	0	2	6	2	2
k	0	0	0	0	0	0	0	0	2	2	6	2
l	0	0	0	0	0	0	0	0	2	2	2	6

Figure 2: The accumulated stimulus co-occurrence matrix for each block in Experiment 1. Each word co-occurred with its intended referent 6 times (A-a, B-b, …) Note that each referent appeared twice with every other referent in its category, but never with referents from other categories. Each 1-to-many label appeared 6 times with each of its intended referents, and 12 times overall.

In the natural category condition, each of the three 1-to-many labels consistently maps to a salient feature present on the stimulus. An additional 12 pairs of testing stimuli were used for a generalization task, using the same category labels that correspond with the stimuli according to their feature.

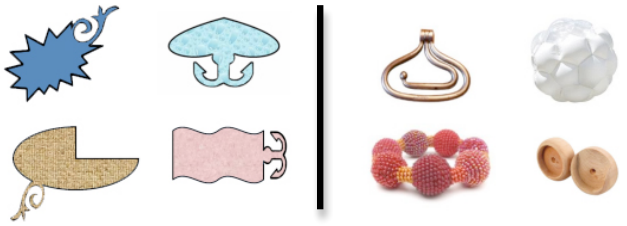

Figure 3: **Left**: In the natural category condition, objects with multiple types of textures and three different protruding shapes were used in training. **Right**: In the arbitrary category condition, objects had no apparent unifying feature.

Procedure

Participants were informed that they would experience a series of trials in which they would hear some words and see some objects. They were also told that their knowledge of which words belong with which objects would be tested at the end. Training for each condition consisted of 36 trials. Each training trial began with the appearance of two objects, which remained visible for the entire trial. After 2 s of initial silence, each word was heard (randomly ordered; 1 s of silence between each word) followed by 2 s of silence, for a total of 9 seconds per trial. After each training block, their knowledge was assessed using 12-alternative forced choice (12AFC) and 3AFC testing: on each test trial a single word was played—a 1-to-1 label or a 1-to-many label—and the participant was asked to choose the appropriate object from a display of all 12 objects (for 1-to-1 labels) or from 3 objects (for 1-to-many labels). For 3AFC testing, one representative from each category was used. The test slides for generalization were the same as the 1-to-many test slides except that the only previously-seen parts of the stimuli were the distinct, protruding shapes (e.g., a hook) that were seen in training to distinguish the different categories. Different stimuli were used in each block. Condition order was fixed.

Results & Discussion

Figure 4 shows the results across all three blocks for each pairing type. Unexpectedly, even in block 1 participants learned a significant number of 1-to-many mappings (M = .49, one-sided $t(32) = 4.95$, $p<.001$, chance=.33) and learned a significant proportion of 1-to-1 mappings (M = .52, one-sided $t(32) = 12.99$, $p<.001$).

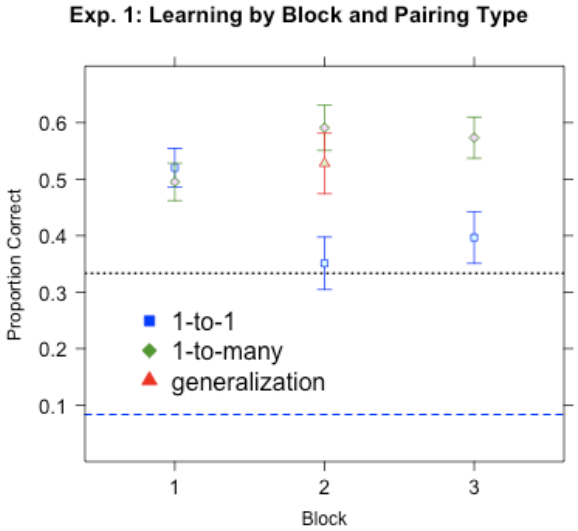

Figure 3: Mean performance for each experimental block by pairing type. Block 1 and 3 were arbitrary groupings and Block 2 was a category grouping; thus, generalization of category type was tested. Error bars show +/-SE. Blue dotted line indicates chance for 1-to-1 learning; black dotted line indicates chance for 1-to-many learning.

After the introduction of a unifying feature, learning of 1-to-1 pairings in block 2 decreased relative to block 1 (M = .35, paired $t(32) = 3.07$, $p<.01$). The perceptual similarity of category members in block 2 may have caused participants to focus on learning 1-to-many mappings, and thus drew attention away from 1-to-1 mappings. In addition, their ability to apply the superordinate name to new referents was reflected in their significantly above-chance (.33) performance on a generalization task (M = .53, one-sided $t(32) = 3.78$, $p<.001$).

Presented with a second arbitrary category condition in block 3, learning of 1-to-1 pairings was significantly lower compared to block 1 (paired $t(32) = 2.96$, $p<.01$), but performance on 1-to-many testing remained higher (M = .57, paired $t(32) = 6.69$, $p<.001$). That is, following the natural category condition in block 2, participants continued to focus on 1-to-many mappings, but still learned 1-to-1 mappings at a proportion over three times chance.

Overall, participants showed evidence of learning both 1-to-1 and 1-to-many mappings in every condition—even in the first condition, when they had no instructions telling them what type of relations would be present, and the referents belonging to each 1-to-many relation (i.e., an arbitrary category) had no unifying perceptual features. Moreover, we observed a shift in learning from block 1 to block 3: after the perceptually-similar category referents of block 2, participants learned more 1-to-many pairings in block 3 than block 1, and fewer 1-to-1 pairings. In Experiment 2, we investigate whether learners can still simultaneously acquire both 1-to-1 and 1-to-many mappings in a still more complex learning situation.

Experiment 2

Experiment 1 showed that humans can simultaneously learn superordinate and basic level names for referents. On each trial, there were two basic level names (1-to-1) and one superordinate level name (1-to-many). Thus, the mutual exclusivity constraint was relaxed and complex relations were formed, with two words referring to each object. After all three conditions in Experiment 1, participants still performed significantly above chance on 1-to-1 associations as well as on 1-to-many associations. However, an alternative learning scenario is an environment in which objects from different categories are learned simultaneously. For example, two referents such as an apple and a carrot could be presented. In this case, each referent has its own superordinate level name (fruit and vegetable, respectively). The learner would need to learn both the superordinate label and basic name label for each object while needing to assign each term to its appropriate referent. The potential for error is much greater because the learner is presented with a more ambiguous learning situation than in Experiment 1, where the superordinate label refers to both displayed referents. Experiment 2 thus presents learners with a four word and two referents (*i.e.* 4x2) on each trial, where two words are category labels referring to a single referent each, and two words are subordinate level names corresponding to one referent each (see Figure 5).

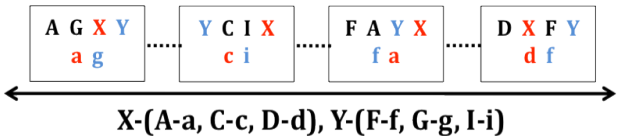

Figure 5: Participants are given 1-to-1 and 1-to-many mappings (e.g. A-a, C-c and X-{a,c,d}) in the context of 4 words and 2 referents per trial.

This extension of the cross-situational paradigm provides additional ambiguity beyond Experiment 1: presented with two more labels than referents on each trial, participants must now learn that the more frequent labels are superordinate, and apply not only to one of the objects on that trial, but also to three other objects seen on other trials. However, given the above-chance performance and particularly exceptional 1-to-many learning, participants may be able to tune themselves into the ambiguous superordinate label to referent pairings after the natural category condition in a manner similar to participants in Experiment 1.

Subjects

Participants were 24 undergraduates at Indiana University who received course credit for participating. None had participated in other cross-situational experiments, including the previous experiment.

Stimuli & Procedure

During training, two objects were shown on a computer screen with four spoken words played sequentially upon presentation of the objects, with time per word equal to that of Experiment 1. New sets of words and referents were used for this experiment. Training for each condition consisted of 36 trials, each lasting 12 s. due to the addition of a spoken category label. Immediately after training for each block, participants were tested for knowledge of the 1-to-1 relations using 12AFC and 1-to-many relations using 3AFC as in Experiment 1. Generalization was also tested for the natural category stimuli. Condition order was fixed.

Results & Discussion

Figure 6 shows results across all three blocks for each pairing type. In Block 1 with arbitrary category referents, participants learned only 1-to-1 names (M = .50; one-sided $t(23) = 7.76$, $p<.001$) while 1-to-many performance was at chance (M = .39, one-sided $t(23) = 1.76$, $p>.05$). Unlike in Experiment 1, block 2 did not see a performance shift. While performance was significant in learning 1-to-1 (M = .46; one-sided $t(23) = 6.45$, $p<.001$) associations, 1-to-many associations were still difficult to acquire, and were not learned significantly above chance (M = .42; one-sided $t(23)$ = 1.85, $p>.05$). Participants may have still not surmised that there was categorical structure involved due to the

confusion of four words per trial (including two category labels). Performance on the generalization task was also at chance, confirming that participants had not yet ascertained the presence and structure of the 1-to-many mappings.

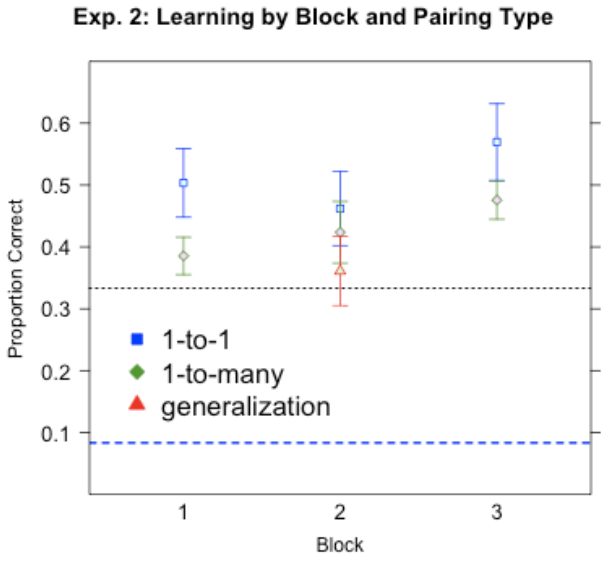

Figure 6: Mean performance by pairing type for each block. Error bars show +/-SE. Dotted lines indicate chance: blue for 1-to-1 pairings (.08); black for 1-to-many pairings (.33).

However, block 3 performance was significantly above chance for both 1-to-1 (M = .57; one-sided $t(23)$ = 7.98, $p<.001$) and 1-to-many (M = .46, one-sided $t(23)$, $p<.001$) associations. Thus, although the higher degree of ambiguity in Experiment 2 made participants take longer to catch on to the presence of multiple superordinate labels on each trial, in the final block they were able to learn these 1-to-many relationships in addition to the 1-to-1 relationships. In comparison to block 3 of Experiment 1, 1-to-many learning in Experiment 2 was significantly lower (Welch's t(55.0) = 2.08, $p<.05$), showing that the superordinate label structure (2 per trial) in Experiment 2 was indeed harder than the structure (1 superordinate label per trial) in Experiment 1.

However, even when participants were uncertain about the meaning of the superordinate labels in blocks 1 and 2, they learned a significant number of 1-to-1 mappings. In block 3 performance, not only did participants learn a significant number 1-to-many mappings, they also learned more 1-to-1 mappings than in the previous two blocks (block 2: paired $t(23)$ = 2.44, $p<.05$, block 1: paired $t(23)$ = 2.03, p=.05). The natural category condition once again provided a clue as to what learning strategy participants need to utilize. However, the significantly lower performance for block 1 in Experiment 2 as compared to Experiment 1 may also indicate interference due to confusion over the two extra labels.

In both experiments, it is important to note that since both 1-to-1 and 1-to-many word-referent mappings learned involving the same referents, each referent was thus part of a 2-to-1 word-referent mapping. Thus, it is possible to determine whether participants learned mappings that violate mutual exclusivity. In both experiments, participants were tested on each referent twice: for the 1-to-1 label (chance=1/12) and 1-to-many label (chance=1/3). Thus, learning that respects ME occurred when participants learn either 1-to-many or 1-to-1 mappings, but not both, and learning that violates ME occurred when participants learn both. As shown in Figure 7, across both experiments and in every block, the average participant learned a significant number of pairings that violate ME as they learned both 1-to-1 and 1-to-many mappings.

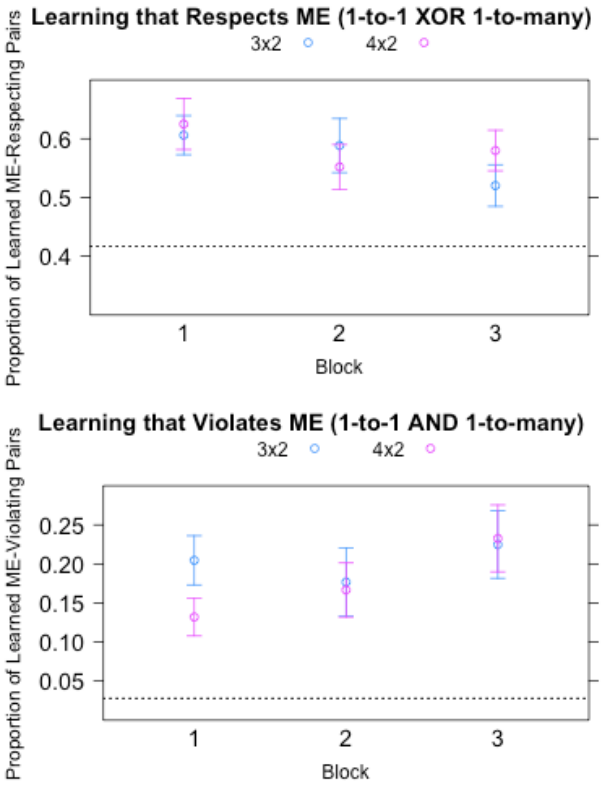

Figure 7: Comparison of proportion of learned ME violating vs. respecting pairs by block for each experiment. Chance (dotted line): Respects=1/3+1/12; Violates=1/3•1/12=.03

General Discussion

While the mutual exclusivity constraint can be a powerful tool for learning 1-to-1 mappings, the hierarchical structure of the real world—which is reflected in natural language—requires people to learn word-referent mappings that are not mutually exclusive. The present study demonstrates that learners learn both 1-to-1 and 1-to-many mappings from situations in which these regularities are simultaneously present.

By the end (block 3) of both experiments, performance for both 1-to-1 and 1-to-many testing was significantly above chance. Experiment 1 shows that participants on

average performed strongly on 1-to-many associations, particularly after the introduction of within-category perceptual similarity in block 2. This may be due to the natural stimuli serving as a primer for learning 1-to-many mappings in block 3. However, although there appears to be a trade-off in learning both types of relationships, participants nevertheless managed to learn both superordinate and basic level names in the first block of Experiment 1. Experiment 2 showed participants could not only learn superordinate and basic level names but can also handle an additional layer of ambiguity when the two referents on a trial belonged to two different superordinate categories. Consistent with Experiment 1, an increase in 1-to-many performance was seen after block 2 was observed in Experiment 2, but 1-to-many performance was overall lower than in Experiment 1. Correspondingly, generalization of superordinate labels to novel objects was also difficult for learners. The more complicated structure (four labels and two referents per trial, representing two categories) in Experiment 2 produces many more possible pairings per trial for a learner to consider. Naturalistic learning situations are even more complex, with multiple co-occurring words, events, and objects (Hart & Risley, 1995); Experiment 2 simulates a more natural scenario in which multiple referents with vague relationships to their superordinate labels are presented. This suggests that infant learning of higher order relations could be guided by creating more unambiguous learning scenarios in order to reduce the likelihood of attribution error.

Interestingly, participants were equally likely to know the superordinate level names (*e.g.,* fruit) regardless of their performance learning basic level names *(e.g.,* apple). Is this due to the mutual exclusivity constraint? In the 3x2 design of Experiment 1, participants were more likely to form a 1-to-many relationship if they do not know the superordinate level name than if they know both (P(Know Superordinate Name | Not Know Basic Name) = .31; P(Know Superordinate Name | Know Basic Name) = .19). The same relationship held in the 4x2 design (.25, .18 respectively). Therefore, participants seemed to form superordinate level relationships more easily rather than basic level relationships.

While the ME constraint may be useful in learning 1-to-1 relationships, the present study's experiments show that participants will focus on forming 1-to-many relationships rather than 1-to-1 relationships if the need to learn higher order relationships becomes apparent, which is often the case in category learning. The strong performance in 1-to-many learning independent of 1-to-1 performance may indicate that people are particularly tuned to learning complex relationships. Every day, we use categories as functional filters of our world to constrain the amount of information we must process at lower (basic) levels (Goldstone & Kersten, 2003). Furthermore, the addition of an exemplar to a category gives us more information about other novel candidate members of the category, allowing learners to generalize as demonstrated in Experiment 1. In future work, we hope to replicate our findings in infants as well as focus on what learning strategies are used by both infants and adults. We also expect that these findings will be useful in constraining formal models of cross-situational word learning.

Acknowledgements

This research was supported by National Institute of Health Grant R01HD056029.

References

Goldstone & Kersten (2003). Concepts and Categorization. In *Comprehensive handbook of psychology*. New Jersey: Wiley; pp. 599-621.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Ichinco, D., Frank, M., & Saxe, R. (2009). Cross-situational Word Learning Respects Mutual Exclusivity. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Frequency and Contextual Diversity Effects in Cross-Situational Word Learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Klein, K. A., Yu, C., & Shiffrin, R. M. (2008). Prior knowledge bootstraps cross-situational learning. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 1930-5. Austin, TX: Cognitive Science Society.

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

Yu, C. & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18, 414-420.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 715–720. Austin, TX: Cognitive Science Society.