Re-learning labeled categories reveals structured representations

Andrew T. Hendrickson; George Kachergis; Caitlin M. Fausey; Robert L. Goldstone
({athendri, gkacherg, cfausey, rgoldsto} @indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 East Tenth Street, Bloomington, IN 47405 USA

Abstract

How do people learn to group and re-group objects into labeled categories? In this paper, we examine mechanisms that guide how people re-represent categories. In two experiments, we examine what is easy and what is hard to relearn as people update their knowledge about labeled groups of objects. In Study 1, we test how people learn and re-learn to group objects that share no perceptual features. Data suggest that people easily learn to re-label objects when the category structure remains the same. In Study 2, we test whether more general types of labeling conventions -- words that do or do not correspond with object similarities -- influence learning and re-learning. Data suggest that people are able to learn either kind of convention and may have trouble switching between them when re-structuring their knowledge. Implications for category learning, second language acquisition and updating representations are discussed.

Keywords: categories, labels, learning and transfer, knowledge change

Introduction

An eighth-grade science student will happily tell you that she has just learned the electrons and nucleus of an atom are very similar to the planets and sun in our solar system. She now has a "multi-body orbiting systems" category with both of these systems as members. This approach is a classic example of transferring knowledge: information she previously learned about the solar system can be applied to the atom. But the story doesn't end there: a physicist will tell you the truth is that an atom doesn't really work that way, electrons are quantum wave functions that do not orbit a central point. New information causes the analogy to break down and if the eighth-grader continues to study physics she will learn new categorizations because the solar system belongs with systems explained by Newtonian mechanics while the atom belongs with quantum systems.

The process of relearning categories and reshaping knowledge is important not only for shifting from novice to expert but also for learning new ways to use words. Many languages group objects, relations and events in different ways (e.g., Majid, Boster, & Bowerman, 2008; Malt, Sloman, & Gennari, 2003; Wolff, Jeon, & Yu, 2009) and second language learners must categorize in new ways in order to speak their second language conventionally.

In laboratory category learning tasks, relearning has been studied using highly structured, binary dimensions. Using 3-dimensional binary stimuli, Kruschke (1996) found a clear hierarchy in the speed of relearning. Preserving the same classification rule but reversing the response options is relearned the fastest, and switching to a new classification rule that involves a previously relevant dimension is more quickly learned than a rule using a previously irrelevant dimension. This pattern of results is consistent with the reversal learning literature (e.g., Kendler & D'Amato, 1954) and successfully modeled with straightforward extensions of many attention-shifting exemplar-based categorization models (e.g., Kruschke, 1996).

The main goal of this work is to examine category re-learning for categories that are not clearly defined by rules. Many categories in the world and in language do not obey a simple rule-based classification scheme, and it is interesting to consider the challenges that learners may face as they structure and re-structure their knowledge about these categories. In this work, we focus on how people learn to group and re-group objects into labeled categories.

Examining how people re-learn to group objects into categories has the potential to reveal insights into three important issues. First, does existing evidence about category learning and re-learning (e.g., with rule-based categories) extend to other kinds of categories? This is important to understand for more general theories of knowledge development. Second, how do people represent labeled groups of objects? What role does a label play in structuring knowledge (e.g., label-as-feature vs. label-as-category-marker, Deng & Sloutsky, 2012; Gelman & Markman, 1986)? This hotly-debated issue may benefit from data about patterns of re-learning because transfer paradigms are a useful way to assess what has been represented, on the logic that people are better able to re-learn structures that closely match what they had originally represented (e.g., Kruschke, 1996). Finally, second language learners are a large part of the world's population and when people learn a new language they often must learn to categorize objects in different ways. What are the mechanisms that guide this re-representation?

As a first step toward these aims, here we consider a variety of potential changes between initial learning and relearning. In the experiments described below, people first learn a category structure and then are asked to relearn across a variety of potential changes. These changes reflect potential real-world relearning situations and the question becomes: what is easy and what is hard to re-learn as people re-structure their knowledge about labeled groups of objects?

Study 1: Re-learning with unrelated objects

In Study 1, we examined several learning and re-learning relationships. In all scenarios, participants first learned to label nine objects. These nine objects were grouped into three categories; three labels were paired with three objects. After learning the initial categorization, participants re-learn. We manipulated what information remained constant and what changed between learning phases. Participants faced re-learning scenarios in which either the objects, the labels, the grouping of objects, and/or the mapping between groups and labels changed (see Table 1). Scenarios in which people quickly adapt to new object-label mappings are likely to be scenarios that conserve whatever representations they learned from labeling the original nine objects. Scenarios that people find more difficult likely overlap less with the learned representations. Thus, how easy it is to adapt to new object-label mappings has the potential to reveal some aspects of how people represent labeled objects.

Several patterns of re-learning data would be informative. In particular, the data may distinguish whether learners represent the information in the following ways: (a) Learners associate each distinct object with its appropriate label, or (b) Learners associate objects with labels, and also represent that the three similarly labeled objects are related to each other. That is, despite the fact that objects never appear together participants may learn the object groupings by virtue of sharing a label. Relearning patterns can also disambiguate between the potential interference or benefit of relearning new pairings or groupings of old objects compared to novel objects. We examine these questions in the following experiment.

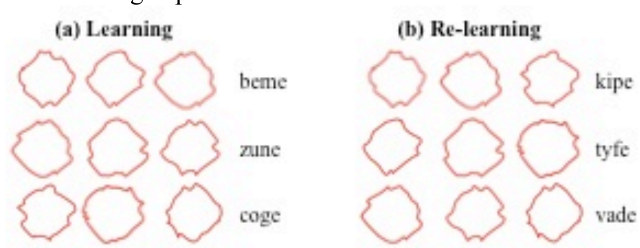

Figure 1: (a) An initial learning structure with 9 objects randomly assigned to 3 labels (by row). (b) A re-learning structure in the Re-Group & Re-name condition. Old objects grouped in a different way with novel labels.

Method

Participants. 180 Indiana University undergraduates participated for course credit. 38 participants failed to complete the study in the allotted time and were excluded from analyses. Participants were randomly assigned to one of seven conditions (n = 19 to 23 for each condition).

Materials. Figure 1 shows the objects and labels that one participant might see during (a) learning and (b) re-learning. For object stimuli, 72 unique segments were formed by fitting a spline through 8 randomly perturbed points along a 90-degree arc. Objects were created by combining 4 segments; a unique set of 18 objects was created for each participant by sampling without replacement from the set of segments and arranging them to form a continuous outline. Labels were novel words from the set of {*beme*, *vade*, *kipe*, *coge*, *zune*, *tyfe*}. Other than the final "e" for all words, no letter was repeated across labels.

Design. All participants completed an original learning phase and then a re-learning phase. For each participant, nine of 18 objects and three of six labels were randomly selected for the initial learning phase. The remaining objects and labels were used in re-learning if needed. Conditions were defined by the changes between learning and re-learning (see Table 1).

Table 1: Re-learning Conditions in Study 1

Re-learning Condition	Items		Grouping Structure Conserved		Object-Label Associations Conserved	
	Objects	Labels	Yes	No	Yes	No
Learned	Old	Old	✔		✔	
Re-map	Old	Old	✔			✔
Re-name	Old	New	✔			✔
Re-group	Old	Old		✔		✔
Re-group & Re-name	Old	New		✔		✔
Recycled Name	New	Old		✔		✔
Novel	New	New		✔		✔

In some cases, new mappings between objects and labels were prompted only by substituting either new objects or new labels for the old objects and labels already learned. For example, a participant in the **Recycled name** condition might have used "*beme*" to label three objects during learning and then learned to re-use "*beme*" to label a novel three objects during re-learning. Similarly, a participant in the **Re-name** condition might have first learned that three objects were all called "*beme*" and then re-learned that they are all called "*zune,*" a name that had not been presented before. In these conditions, participants saw one set of objects [or labels] during learning and a different set of objects [or labels] during re-learning.

In other cases, new mappings between objects and labels *reorganized* the structure of previously learned categories using old objects. For example, consider a participant who first learned that three objects are each called "*beme,*" another three are each called "*zune,*" and a final three are each called "*coge.*" In **Re-map**, they might later re-learn that the first three are now called "*zune,*" the second three "*coge*," and the third three "*beme*." The grouping of objects is intact, the mapping between objects and labels is not. In **Re-group**, they might later re-learn that "*beme*" is now used for one object previously called "*beme,*" one object previously called "*coge,*" and one object previously called "*zune.*" This breaks the grouping of objects. In **Re-group & Re-name**, they might later learn to use "*kipe,*" a label not previously presented, to label three objects that had previously been called "*beme,*" "*coge,*" and "*zune.*" This breaks the grouping of objects and uses new labels.

Two other conditions gave participants no conflict with what was learned before – in one they continued with the same objects, labels, and mapping between them as they had originally learned (**Learned**) and in another they learned about totally new objects and labels (**Novel**).

Procedure. On every trial, participants saw one object and three possible labels (Figure 2). The starting position of the cursor was equidistant from all response options and the location of labels was randomly determined on every trial. Participants were asked "*Which category does this belong in?*" The object and labels remained on the screen until the participant made a response. Afterward, feedback appeared above the object for 1200 ms: "*Correct [Incorrect]! This is a* ______." Feedback included the correct label for all responses. There was a 400 ms pause between trials.

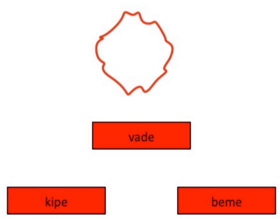

Figure 2: A sample trial.

Participants completed the learning phase in blocks of 9 trials. In each block, every object was presented once. Participants continued in this learning phase until they were correct on at least 8 out of 9 trials in a block for 4 consecutive blocks. Thus, everyone learned the original categories to criterion.

Participants then started the re-learning phase, reading these instructions: "*You are doing great. In the next section the categories may change*." The re-learning phase consisted of 5 blocks in which all of the 9 stimuli were presented 3 times.

Results

Initial Learning. The minimum number of blocks to reach criterion during the initial learning was 7 and the maximum number was 94. Participants reached criterion with a mean of 34.1 and a median of 31.0 blocks.

Re-learning. People's performance in the relearning phase depended on block and condition. Re-learning data were analyzed using an ANCOVA, with categorization accuracy as the dependent variable, condition (7 levels) as a factor, and block (5 blocks) as a covariate. There were main effects of Condition ($F(6, 135) = 23.1$, $p < 0.0001$) and Block ($F(1,561) = 491.5$, $p < 0.0001$), and a significant interaction between Condition and Block ($F(6,561) = 12.9$, $p < 0.0001$).

In order to interpret the interaction between Condition and Block, the trajectory of accuracy across block for each subject was clustered into groups and the distribution of each category within clusters was compared. The trajectories were clustered with methods to estimate the appropriate number of clusters (Kaufman & Rousseeuw, 1990). This technique identified two clusters, where one cluster was characteristic of three conditions: Learned (all 19 subjects), Re-name (19 of 20), and Re-map (16 of 21). These were conditions in which the original grouping of

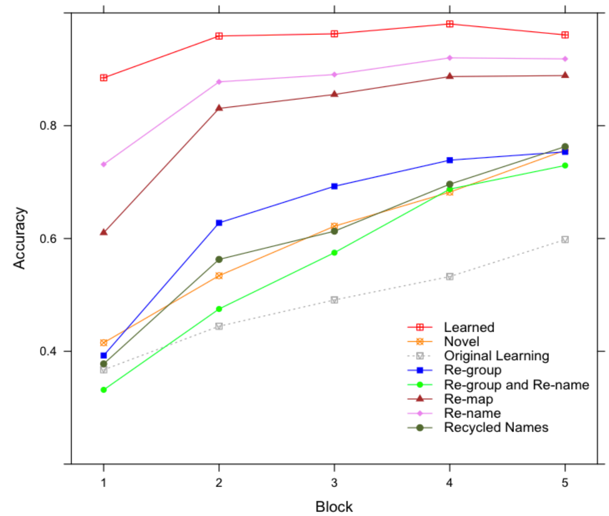

Figure 3: Participants' mean accuracy by condition across re-learning in Study 1. Re-learning was faster than Original Learning. Of the re-learning conditions, Learned, Re-name, and Re-map show higher performance than the others.

objects were conserved from learning to re-learning (i.e., the first three rows of Table 1). The other cluster was characteristic of three conditions: Novel (14 of 19), Recycled Name (14 of 20), and Re-group & Re-name (17 of 23). The Re-group condition was equally split between both clusters (10 of 20 in each).

Discussion

Performance during relearning depended on whether or not the groupings of objects were maintained from learning to re-learning. The three conditions in the higher accuracy cluster – Learned, Re-map, and Re-name – all conserved the grouping of objects from learning to relearning. The conditions strongly consistent with the lower accuracy cluster – Re-Group & Rename, Recycled Name, and Novel – did not preserve any groupings, either because they consisted of novel objects or because the three objects mapped to one label in relearning had not been mapped to a common label during learning.

The participants in the Re-group condition were equally split between the two clusters. In this condition, the nine objects that people saw during learning and re-learning were the same. Three of these were associated with the same name (one per label) during learning and re-learning, while the remaining six became associated with a different name during relearning (due to re-grouping the objects). The split between clusters suggests participants in this condition may have used different strategies, with some individuals less disrupted by the re-grouping because they were able to use object-label mappings that were identical between learning and re-learning but others learned representations that strongly relied on groupings, which were not conserved.

Study 1 suggests that labels — novel or re-purposed — are relatively easy to re-map to a group of objects. Learning new object groupings or re-groupings is more difficult.

These data suggest that learners may have successfully represented more than object-to-label mappings for each individual object and instead capitalized on structure among objects. We discuss this idea further after considering another interesting case of learning and re-learning.

Study 2: Re-learning with related objects

The objects in Study 1 could be grouped into categories in only by using the labels because the objects shared no segments. How do people learn and re-learn *related* objects? In what way(s) does similarity among objects influence object grouping and impact re-learning?

To examine these issues, we again used a category learning and re-learning paradigm in Study 2. There are three main differences in Study 2: (a) objects were created such that they had structured similarity because some objects contained the same segments in the same locations; (b) the nature of mapping between objects and labels was manipulated to be consistent or inconsistent with object similarity. The mapping from training could either persist or switch from learning to re-learning; (c) all re-learning scenarios used novel labels and old objects. The non-control conditions in Study 2 all involved breaking the grouping of objects with the goal of understanding how relearning is influenced by the mapping between object similarity and category labels.

Method

Participants. 107 Indiana University undergraduates participated in this study for course credit. 30 participants (evenly distributed across conditions) did not complete the study in the allotted time and were excluded from analyses. Participants were randomly assigned to one of six conditions (n = 11 to 14 per condition).

Materials. Objects were created by combining four segments from a set of 72 segments. Nine objects were created for every participant. Instead of randomly selecting unique segments without replacement for each location of every object, some segments repeated across objects.

Specifically, two segments of every object also appeared in exactly two other objects (see columns 2 and 4 of the stimuli dimensions in Table 2). No two objects shared more than one segment and the location of the repeated segments was constrained so that they were not adjacent to each other. The other two locations were unique segments sampled without replacement (columns 1 and 3).

Further, in order to avoid possible preferences for a particular spatial location (e.g., whatever appears in the "top left quadrant" is easier to learn because of looking tendencies), a random number between 1 and 360 was selected for every participant and all objects for that person were rotated that many degrees.

In all conditions, different labels were used during learning and re-learning. For learning, three labels were randomly sampled without replacement from the same set of labels used in Study 1. The remaining three labels were used during re-learning.

Table 2: Stimuli and condition structure in Study 2

Stimuli	Learning		Re-Learning	
	Many	One	Many	One
1111	A	A	D	D
2122	C	A	E	E
3133	B	A	F	F
4241	B	B	E	D
5252	A	B	F	E
6263	C	B	D	F
7371	C	C	F	D
8382	B	C	E	E
9393	A	C	D	F

Objects: Four segments per object, columns indicates a location. Within a column, the same number indicates an identical segment. Across columns, numbers are unrelated. *Learning & Relearning:* Letters indicate labels. In "One" columns, a label matches one segment (Learning-One, Segment 2; Relearning-One, Segment 4). In "Many" columns, no single feature matches a label.

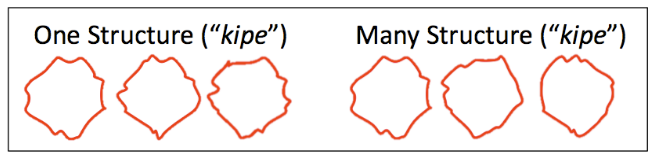

Figure 4: Example categories (Study 2). "One": shared segment in upper right. "Many": All unique segments.

Design. Categories were defined by the nature of segment-to-label mapping. The One category structures had a one-to-one mapping between segments in a location and labels (see Table 2). The Many structures did not have a one-to-one mapping, and never shared a segment within a category.

Conditions were defined by the type of structures in learning and re-learning. Learning and re-learning structures were combined so that some participants continued in the same style of mapping (One-One; Many-Many conditions) while others switched (One-Many; Many-One conditions). All participants in these conditions learned new ways to group objects during re-learning regardless of the type of mapping in learning and re-learning, the group of objects that shared the same label changed.

Two additional control conditions did not change the category structure between learning and re-learning though novel labels were introduced. In "Same One" participants started and remained in a "One" structure and in "Same Many" participants stayed in a "Many" structure.

Procedure. Learning and re-learning was exactly like Study 1. After re-learning, people did a re-test in which they tried to recall the label that they had *originally* learned for every object. They were instructed "In this final section we will ask you about the first labels you learned. You will not be told if you were correct or not, please do the best you can." The re-test trials only displayed the original labels and no feedback was provided. People saw each object 8 times.

Results

Initial Learning. People reached criterion in about the same number of blocks during each kind of initial training (One: $M = 35.2$, Many: $M = 41.7$, $t(75) = 1.5$, $p = 0.14$).

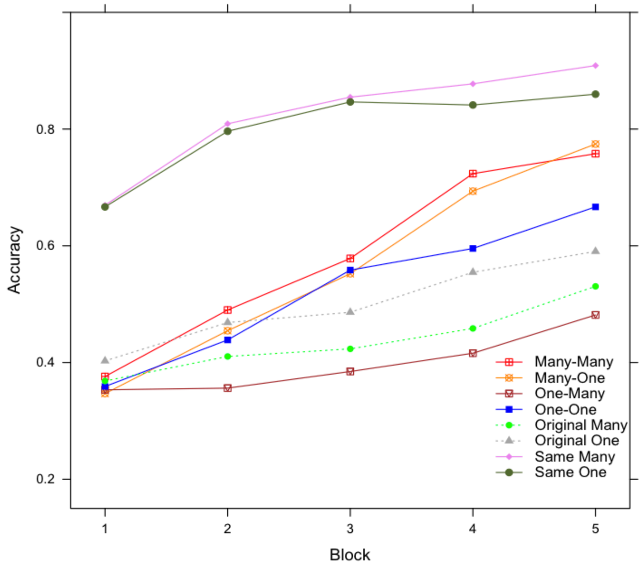

Figure 5: Participants' mean accuracy in Re-learning (Study 2). The Same conditions are easy, followed by Many-Many, Many-One, and to some extent One-One. One-Many is more difficult, like the Original Learning.

Relearning. People's performance in the relearning phase depended on block and condition (Figure 5). Re-learning data were analyzed using an ANCOVA, with accuracy as the dependent variable, Condition (6 levels) as a factor and Block (5 blocks) as a covariate. There were main effects of Condition ($F(5,71) = 12.7$, $p < 0.0001$) and Block ($F(1,302) = 275.4$, $p < 0.0001$), and a significant interaction between Condition and Block ($F(5, 302) = 9.5$, $p < 0.0001$).

Like Study 1, a clustering analysis was performed using the accuracy data across block for every participant and was best fit with 3 clusters. The highest accuracy cluster was most consistent with the two conditions in which object groupings did not change from learning to re-learning: Same Many (10 of 12) and Same One (12 of 14). A second cluster was most consistent with the two conditions that did change groupings but started with the Many mapping: Many-One (7 of 11) and Many-Many (9 of 12). The third cluster was most consistent with the One-Many condition (10 of 13). The One-One condition was equally consistent with the second and third clusters (6 of 13 in each). Thus, switching from a One structure, where a segment and label are paired, to a Many structure, where many segments map to a label, appeared to be particularly challenging.

Re-test. After re-learning, people's memory for the originally learned labels did not appear to be strongly influenced by Condition or Block. Memory accuracy was analyzed using an ANCOVA with accuracy in the re-test as the dependent measure, Condition (6 levels) as a factor and Block (4 blocks) as a covariate. A marginal effect of Condition was found ($F(5,70) = 2.2$, $p = 0.062$). No significant effect of Block was found ($F(1,222) < 1$, $p = 0.5$). There was no significant interaction between Condition and Block ($F(5, 222) = 1.4$, $p = 0.21$).

To better understand the marginal effect of Condition, pair-wise post-hoc tests showed that the accuracy in the Many-Many condition ($M = 0.73$) was significantly lower than Same-Many ($M = 0.91$, $p = 0.006$). All other conditions were not significantly different ($p > 0.01$).

Discussion

When people learned novel labels for objects, they had the easiest time if the original groupings of objects were preserved. Using new words to talk about old groups was equally easy whether or not the original groupings were mapped to labels based on repeated, shared segments or based on individual items. If the grouping of objects was disrupted, however, people who had originally learned a Many mapping (i.e., item-specific association with labels) re-learned faster than those who had learned a One mapping (i.e., repeated-segment association with labels).

Why is it easier to learn new labels for new groups after having first learned an item-specific labeling convention than after having learned a labeling convention that capitalizes on perceptual similarity? Learning that words are used for similar objects (i.e., One-One and One-Many) may have directed attention to a common, predictive feature for each word. Re-learning to label these objects may be hard when this feature is no longer predictive. But shifting attention to different features might be easier if you have learned that words are used on an object-by-object basis – this is a useful approach during "re-learning" even when the particular mappings change. Thus, generalizations about "labeling conventions" may influence later re-learning.

Interestingly, people may learn labeling conventions that capitalize on perceptual similarity in different ways. In this study, participants in the One-One condition were equally split between the two clusters. The group of people who were clustered with the Many-Many and Many-One conditions may have been able to re-learn easily because they discovered the new shared feature during re-learning. In addition to the original associations between objects and labels, these people may have learned that there *is* a feature that predicts a label. They then successfully transferred this generalization during re-learning. The other group of people in this One-One condition may have learned the original association between a shared feature and its label but not the higher order pattern and could not later shift their attention.

Re-learning is even more challenging if people must overcome not only a learned association between a specific feature and a label, but also the higher-order generalization that there *will be* one feature that predicts each label. This is the situation of learners in the One-Many condition, and indeed, they performed the worst during re-learning.

Is there more than one way to successfully re-learn after having learned a labeling convention that is item-specific? This is an open question. In the Many-Many and Many-One conditions, people initially learned item-specific mappings between objects and labels. In the re-learning, even though people in the "Many-One" condition could have learned to associate labels with shared features, they also could have learned these associations in an item-specific way. The fact that people in the Many-Many and Many-One conditions showed similar re-learning trajectories suggests that this may have been a common strategy. Thus, it is possible people in all re-learning scenarios were likely to continue using whatever labeling convention they had originally used (either item-specific, or shared feature). It just happens to be that an item-specific strategy leads to success throughout learning and re-learning, while a labeling convention that relies on perceptual similarity (shared features) does not.

General Discussion

Data from two experiments suggest that people easily learn to re-label objects when the category structure remains the same from original to subsequent learning. Further, people can learn multiple labeling conventions – words do or do not correspond with object similarities – but may have trouble switching between them when restructuring their knowledge. Like advantages for using relevant dimensions when re-learning rule-based categories (Kruschke, 1996), there are advantages to using existing category structures when re-learning arbitrary or similarity-based categories.

The advantage for shared structure, as well as shared labeling conventions, may help to explain some difficulties that adults face when learning a second language. It should be especially hard to learn new labels for objects that are organized differently in the two languages compared to objects that are categorized similarly. Moreover, an intriguing possibility is that once learners have made higher-order generalizations about the kinds of non-linguistic structure that predict labels, they may find it hard to "start from scratch" and build new associations as they construct different similarity spaces. Interesting test cases of this idea would be to see if L2 learners make systematic labeling errors based on L1 structure (and change over the course of L2 learning), and also whether object similarity spaces are predictably different in bilinguals than monolinguals.

Labels may play a critical role in forming categories and shaping representations (e.g., Goldstone, 1994; Goldstone & Hendrickson, 2009) and these representational changes may persist during subsequent use. The re-learning tasks used here raise interesting questions about how labels may wax and wane as drivers of representation and re-representation. We suggest that a focus on *change over time,* together with approaches that test knowledge structures at a single moment will enrich our understanding of these issues.

Learning any particular structure can be a double-edged sword. Subsequently learning a very similar structure may be easy, but it may be much harder to learn very different structures. In the case of language, different labeling conventions may promote relatively richer or shallower encoding of individual object representations. For example, if it is possible to successfully label objects by paying attention to a single shared feature, then using labels to learn these categories may promote a representation that prioritizes this feature. But, if it is necessary to pay attention to multiple aspects of objects in order to talk about them, then people will (e.g., Slobin, 2003). Thus, the complexity of the mapping between labels and structures might shape subsequent representations via perceptual learning (Goldstone & Hendrickson, 2009). It may be easy to use them in some situations, but harder in others.

In general, transferring knowledge to new situations is easier after deeply encoding the relevant structure and certain kinds of initial exposure promote this kind of encoding (e.g., Day & Golstone, 2011; Gentner & Markman, 1997). It is an interesting open question to consider what aspects of language learning – what labeling conventions, used at what points of learning about the relevant structures and categories in one's world – support different trajectories of developing object representations.

Acknowledgements

This research was supported in part by National Science Foundation REESE grant 0910218, Department of Education IES grant R305A1100060, and National Institute of Health and National Research Service Award HD007475-17.

References

Day, S., & Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Deng, W., & Sloutsky, V.M., (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science.*

Gelman, S.A., & Markman, E. (1986). Categories and induction in young children. *Cognition*.

Gentner, D., & Markman, A. B. (1997). Structural alignment in analogy and similarity. *American Psychologist*.

Goldstone, R .L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*.

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical Perception. *Interdisciplinary Reviews: Cognitive Science*.

Kaufman, L., & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Kendler, & D'Amato, (1954). A comparison of reversal shifts and non-reversal shifts in human concept formation behavior. *Journal of Experimental Psychology*.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*.

Majid, A., Boster, J.S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*.

Malt, B.C., Sloman, S.A., & Gennari, S.P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*.

Slobin, D.I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadon (Eds.), *Language in mind: Advances in the investigation of language and thought*. MIT Press.

Wolff, P., Jeon, G., & Yu, L. (2009). Causal agents in English, Korean and Chinese: The role of internal and external causation. *Language and Cognition*.