Exploring the Role of Verbal Category Labels in Flexible Cognition

Jackson Tolins (jtolins@usc.edu)
Department of Psychology, 1156 High St.
Santa Cruz, CA 95064 USA

Eliana Colunga (colunga@psych.colorado.edu)
Department of Psychology and Neuroscience, 345 UCB
Boulder, CO 80309 USA

Abstract

Research under the paradigm of the label feedback hypothesis has proposed a causal role for verbal labels in the online learning and processing of categories. Labeled categories are learned faster, and are subsequently more robust. The present study extends this research paradigm by considering the relationship between verbal labels and flexible categorization. Flexibility is a key trait of human cognition, and flexible categorization is important in a number of tasks. Participants learned to categorize ‘friendly’ and ‘unfriendly’ aliens either with or without names, followed by a transfer task. While selective attention to a particular dimension slowed relearning, no effect of label was found for either category learning or relearning with one exception; labels facilitated flexibility when selective attention was not involved in the transfer. The inability to replicate effects of verbal labels in category learning using similar methodologies raises interesting theoretical issues, questioning the extent to which this relationship applies.

Keywords: Categorization; Label Feedback Hypothesis; Flexible Cognition; Selective Attention

Introduction

Language, along with use in communication, provides a symbolic system of representation through which a speaker contemplates the world around them. The emergence of the capacity for symbolic representation transformed human cognition (Deacon, 1997; DeLoache, 2004), permitting abstract thought and making possible cultural transmission of knowledge. Yet the relationship between language and other cognitive processes is still controversial. For many who view language as a distinct mental module (Gleitman & Papafragou, 2005; Pinker, 1995), language is merely a formal medium that is used to describe mental representations, while remaining independent of the concepts they express (Li & Gleitman, 2002). Recent work in understanding the relationship between language and thought has provided evidence against this disassociation. Instead, it has been suggested that language is best understood as built upon domain general cognitive processes, and thus potentially in a mutually transformative relationship with these processes (Bowerman & Choi, 2001; Gumperz & Levinson, 1996).

With habitual use of the specific set of conceptual symbolic representations afforded by a language, an individual may be biased towards these representations in problem-solving and other cognitive tasks. How a language may accomplish this is not well understood. One possibility is that language reduces the ability to flexibly adjust categories outside the structure provided for by the words of a particular language. As such, it is important to consider the influence of language on the ability to dynamically activate and modify the cognitive process of categorization in response to changing task demands. The ability to think and act adaptively, while not a uniquely human trait, is a mental capacity uniquely well developed in human cognition and intelligent behavior (Deák, 2003). For the purposes of the current study, flexible cognition will be defined as a property of the cognitive system, rather than a specific mechanism or process (Deák, 2003; Ionescu, 2012). This definition allows for the consideration of flexible cognition in the interaction of interest; that between categorization and language, specifically verbal labels.

Recent work lead by Gary Lupyan and colleagues on the role of labels in categorization has demonstrated a special status afforded to verbal labels (see e.g. Lupyan, Rakison, & McClelland, 2007; Lupyan & Thompson-Schill, 2011). Verbal labels participate in the learning of categories, facilitating learning, creating mental categories that are more robust than when the categories are learned without words (Lupyan, Rakison, & McClelland, 2007), and encouraging selective attention (Brojde, Porter, & Colunga, 2011). However, no study has looked directly at the influence of verbal labels on the perceptual and attentional processes that underlie flexibility after learning. Similarly, while a number of studies have looked at how language aids in an individual’s ability to flexibly adjust the level of categorization, or switch from taxonomic to thematic (Blaye, Bernard-Peyron, Paour, & Bonthoux, 2006), no previous research has investigated how individuals flexibly adjust their categorization strategies in regards to the same domain, on the same level. The present investigation seeks to illuminate further the relationship between verbal labels and the cognitive processes underlying categorization. In developing an understanding of the role that verbal labels play in the construction and maintenance of categories, we further our understanding of the relationship between language and the domain general cognitive processes, such as categorization, upon which language is built.
Background

Flexibility in Categorization

Categorization, the process by which discriminably different things are classified into groups and therefore responded to in kind, is a ubiquitous cognitive operation relevant to all aspects of human life. How categories are learned is a key issue in understanding the relationship between verbal category labels and flexibility in cognition. A number of studies have demonstrated that the relationship between perceptual descriptions, how the category or concept is defined, and conceptual representations, such as verbal labels, are mutually influential (Goldstone, 2000; Lin & Murphys, 1997). It is widely accepted that adults tailor the categories they form to the current demands of the task or situation (Barsalou, 1983), and can spontaneously group objects in several ways (Ross & Murphy, 1999). Categorical flexibility is thus a within-subject variable corresponding to the ability to switch, (or relearn), between different representations of a given object or set of objects.

Related work has focused on the way that categorization influences perceived similarities (e.g. Goldstone, Lippa, & Shiffrin, 2001). According to these studies, conceptual and categorical flexibility must be accompanied by flexibility in perceptual and attentional processes (Goldstone 1998). Two mechanisms are considered key to perceptual category learning and flexibility: selective attention and differentiation of dimensions (Goldstone & Steyvers, 2001).

Selective attention refers to the process by which, in categorization learning, individuals learn to attend to some features of the objects and ignore irrelevant features. Selective attention is key to models of categorization such as Nosofsky’s (1986) exemplar model, in which an object is measured in similarity compared to a stored category member in a multidimensional space. The distances between points along dimensions within this space compress and expand depending on the attention given to particular dimensions. Dimensional differentiation refers to the psychological process by which previously unified dimensions become perceptually and cognitively distinct. For example, in developing categories for circles and squares one must first learn to separate the dimension of shape from task-irrelevant dimensions such as color or size. In order to study these mechanisms, Goldstone & Steyvers (2001) applied a learning/transfer task, wherein subjects first learned to distinguish between two categories, and then at transfer had to relearn the categories based on altered relevance of dimensions. By making dimensions that were previously diagnostic for categorization unimportant, or the reverse, allows for a measure of the role of selective attention in categorical flexibility. Similarly, new dimensions may exist in the transfer stimuli set that did not exist in the training set, allowing a separate measure of dimensional differentiation.

Categorization and Verbal Labels

The processes of selective attention and dimensional differentiation in categorization lead stimuli to be considered more similar when in the same category, and more easily distinguishable when in different categories (Harnad, 1987). Recent studies have demonstrated that verbal labels influence categorization, speeding up the attentional processes that focus in on diagnostic properties of categorized objects. It has been suggested that simply sharing a label, defined as a name for a category, causes two objects to be perceived as more similar than those that do not (Lupyan et al., 2007).

There are a number of explanations for this relationship. Researchers have provided evidence that labels offer more maximally informative feedback during categorization learning, making rule-based categories, those categories that are learned explicitly with diagnostic rules that are easily verbalized (Ashby & Maddox, 2005), easier to learn (Maddox et. al, 2008). Others consider labels as physical, external symbols upon which our categories are hung (Clark, 2006; Lupyan et al., 2007). In this sense, language is viewed as a self-constructed cognitive niche, with words providing the material scaffolding required to promote abstract thought and reason, by providing a target for more basic capacities such as statistical and associative learning (Clark, 2006). These latter theories have been generalized by Lupyan within the Label Feedback Hypothesis framework (Lupyan, 2007).

Labels have been implicated in the learning of categories, but what of their maintenance and adjustment? Lupyan, Rakison, and Mc Clelland (2007) provided evidence that categories associated with verbal labels are not only learned faster, but are maintained more robustly after initially training. If one of the main uses of language is the creation of associations between concepts and words in such a way that the labeled concepts are learned fast and remain more robust, it is possible that a verbal label will also reduce the categorical flexibility by strengthening selective attention to a diagnostic dimension. In contrast, if labels, as suggested by Maddox et al. (2008), simply aid in categorization of rule-based categories by providing a more maximally informative feedback mechanism, it is possible that labels may also positively affect categorical flexibility.

The Current Investigation

The present study seeks to add to the literature on labels and categorization by investigating the rigidity of categorization both with verbal labels and in their absence. When an individual needs to restructure the categorical divisions of a particular domain, especially when this restructuring requires a shift in attention to a previously non-diagnostic dimension, having verbal labels for categories already established could slow down the relearning curve. The influence of verbal labels on learned sensitivity to dimensions was tested using a category-learning paradigm in which participants received an initial category learning followed by a relearning transfer task, in which either the
diagnostic dimension changed, requiring a shift in selective attention, or the behavioral response but not the diagnostic changed (see Transfer Procedure below).

**Methodology**

**Subjects** 192 participants were drawn from the undergraduate psychology subject pool at CU, Boulder, in exchange for course credit. Subjects were randomly assigned to either label or no label training conditions and one of three transfer conditions, giving six total conditions.

**Materials** Categories were organized based on the kind of eyes “aliens” exhibited. To this end, 36 gabor patches were created, varying along the dimensions of frequency and orientation (figure 1); these patches were embedded in the stimuli as the aliens’ eyes.

![Example stimuli demonstrating the range of frequencies across and orientation downward.](image)

**Training Procedure** Following the procedure from Lupyan, Rakison, & McClelland (2007), participants were told that they were to take part in a NASA training program before traveling to a newly found planet. In training, it was explained that previous explorers to the planet had discovered two aquatic alien species, one of which was friendly and could be approached, and one that was dangerous and should be avoided. In the label conditions, the participants were told that the explorers had decided to name the aliens, and that the friendly aliens were named ‘Gowachi’, while the dangerous aliens were named ‘Caleba’. Thus, participants were asked to learn to distinguish between two categories within a set of novel stimuli. This distinction was based on either the orientation or the frequency of the alien’s giant eye. Individual trials began with a fixation marker in the middle of the screen, presented for 500 milliseconds. For each trial, an alien was presented briefly, (500 ms), before a scuba diver appeared in one of four locations; above, below, or on either side of the alien. The participant then decided whether to approach or escape the alien using the directional keys on a standard keyboard. For example, if a scuba diver appeared on the left of a friendly alien, the participant should press the “right” key to move the scuba diver closer. After a response was made, feedback was provided in either minimal (a chime for correct, a buzz for incorrect) or maximal (minimal feedback + correct category label) conditions. If the participant waited for longer than 3 seconds, feedback was given without response. After the feedback, the alien and scuba diver remained on the screen for additional 800 ms before the start of the next trial and the representation of the fixation marker. Each unique alien + diver trial was presented once in random order, for a total of 144 trials of training (36 alien exemplars x 4 diver locations). All subjects received the same number of categorization learning trials and had equal exposure to the stimuli across conditions.

**Transfer Procedure** After training was complete the participants were told that they were now ready to travel to the Planet Teeb. In all but the control, or 0 degree, transfer conditions, upon arrival on the planet the participants were alerted that something has gone wrong, and that the aliens are not behaving as expected. Participants in these conditions faced two distinct relearning tasks. In the 90 degree transfer condition, the diagnostic dimension changed, requiring a modulation in selective attention. Participants who learned during training that the friendly aliens had thick bands in their eyes, and the unfriendly aliens thin ones, here had to learn to categorize the friendly and unfriendly aliens based on the steepness of the orientation of the bands, ignoring thickness. This meant that half of each category learned during the first phase subsequently became part of the new category structure learned during transfer testing, or that half of the Gowachi must now be considered Caleba and the reverse. For the 180 degree transfer condition, the diagnostic dimension remained the same, but the escape/approach responses were switched. Here, participants who first learned that aliens with steeply oriented bands in their eyes were friendly now had to learn to treat them as unfriendly, or that the Gowachi and Caleba were opposite what had been learned. These two transfer conditions were compared to the 0 degree transfer condition, in which no change between the training and testing occurred.

Having all conditions transfer to the same categorization allowed for a clear relationship between initial categorization and participants’ ability to relearn categorization strategies flexibly (see e.g. Goldstone & Steyvers 2001). The post-transfer phase consisted of a second set of 144 randomized trials. During the transfer phase trials only minimal feedback (chime or buzz) were given in all conditions, whether label or no label.

**Results**

Trials were grouped into blocks of 36, giving four blocks each for training and transfer phases. Each correct trial was scored as 1, each incorrect trial as 0, and each trial in which the participant did not answer was dropped. Accuracy across
block was then calculated. The data from those participants who did not reach at least 50% accuracy by the end of training were not included (13 participants in total). Data was then entered into a mixed factor ANOVA.

First, we tested for an effect of label on training, collapsing across transfer type, to see if previous findings on the advantage of having a label would replicate, (Lupyan, Rakison & McClelland, 2007). However, while participants did learn to categorize correctly $F(3, 438) = 103.42$ ($p < \ .001$), there was no main effect of label type on this learning trajectory ($p = .312$). A similar pattern was seen in the testing phase, with a significant effect of block ($F(3,438) = 13.140$ $p < .001$), without an effect of label, or a label by block interaction ($F(3, 438) = 1.263$, $p = .287$), (See Figure 2). There was one significant four-way interaction involving label that will be discussed below.

Turning from label to transfer type, while there was no main effect of transfer type ($F(2, 146) = .104$, $p = .901$), there was a significant interaction between phase (whether training or transfer) and transfer type ($F(2, 146) = 80.553$, $p < .001$), with accuracy worse when transfer required a switch in selective attention, (see figure 3).

Of final interest was a significant four-way interaction between phase, block, label type, and transfer type ($F(6, 438) = 2.18$, $p < .05$). As this was the only significant interaction involving label, this interaction was pursued further, with the analysis first involving separating out each transfer type. For the 0 and 90 degree transfer conditions, there were the expected effects of block and phase (all $ps < .01$), but no main effects or interactions involving label (all $ps > .05$). In the 180 degree transfer condition, however, a significant interaction of phase*block*label type was found, ($F(3, 132) = 4.527$, $p < .05$. Using a general linear model to explore this interaction further, we found that for the first two blocks of transfer in the 180 degree transfer condition, there was an interaction between block and label type ($F(1, 44) = 11.595$, $p < .001$), (see figure 4). Thus, there was evidence for an effect of label on transfer learning in the condition that required not a shift in attention, but a shift in the behavioral response from what had been learned in training.

![Figure 2: Average accuracy by block for label and no label conditions collapsing across transfer type.](image2)

![Figure 4: Average accuracy by block for participants in the label and non-labeled conditions of the 180 degree transfer.](image4)

**Discussion**

The results of the present experiment did not find support for a general advantage for learning categories with labels over categories without labels as seen in previous similar experiments (Lupyan et al., 2007). One important difference between the stimuli used here and that used by Lupyan and colleagues is that their aliens were categorized by shape features, whether the ones in the present study were categorized by what could be seen as textural features. The lack of a label advantage in learning is in line with previous work showing that the effect of labels depends on the sort of categorization being learned (Brojde et al., 2011).

More interesting to the question of this paper, however, is the way labels influenced performance at transfer. In the 180 degree condition, when participants had to relearn that those aliens who had been approachable were now not approachable and vice versa, there was a significantly faster recovery after transfer for those participants who were provided with labels during training. Our results suggest that labels play a positive role in the relearning of
categorization when the boundaries of the categories do not change, and the relevant dimension does not change, but the categorical behavioral responses do, (i.e. whether the astronaut should approach or retreat from the alien). Having verbal labels for the categories allowed the participants to more flexibly adapt to the changing task demands. It is possible that since verbal labels become attached to the categories which they are used to express (Lupyan & Thompson-Schill, 2011), that when the categories themselves do not change, but only the responses change, these labels continue to act as more easily computed symbolic abstractions of the categories for which they stand. It then becomes possible for the participants in the 180 degree transfer condition to switch from 'Gowachi' and 'Caleba' to 'not Gowachi' and 'not Caleba'.

The visibility of this effect of label on transfer flexibility seems to be made possible by the low cost of transfer when the transfer does not involve modulation of selective attention. The cost of transfer, however, was much larger for those who had to relearn their categorization strategies based on a previously unimportant dimension. Those participants who learned during training to categorize based on frequency of the lines of the eyes and discovered on the planet that the aliens were either friendly or unfriendly based on the orientation demonstrated reduced ability to flexibly adjust to this new categorization strategy. While selective attention is an important process in the development of accurate categorization (Goldstone, 1998), it also reduces the degree of flexibility present in categorization cognitive processes.

At transfer, these participants must not only learn to pay more attention to the previously ignored dimension, they must also inhibit attention to the previously diagnostic features (Goldstone & Steyvers, 2001). This is demonstrated by the comparison of the four blocks of training for the identification condition with the four blocks of transfer for those participants whose transfer included a change in the diagnostic dimension, despite having had 144 trials more experience than those approaching the task for the first time. This is a clear indication of the cost that comes with increased attention to one historically predictive dimension combined with decreased attention to all other dimensions. This is in contrast with Goldstone and Steyvers (2001), who found that when the categorization rules are orthogonal, participants do no differently than those learning a completely new set. Their analysis of this finding was to posit an equalizing effect of negative transfer from selective attention with positive transfer from dimensional differentiation, meaning that regardless of the type of transfer, it helps to have practice in separating the two perceptual dimensions of the stimuli. By matching the same categorization strategy across training, taken as a control, and transfer, rather than having participants relearn a completely new category during transfer, we demonstrate that the positive effect of dimensional differentiation is not large enough to make the performance of those participants who transferred across dimensions on par with those coming to the same task without any previous experience.

This role of selective attention in reducing flexibility was not, however, modulated by the presence of verbal labels corresponding to the categories being learned. While participants did learn the correct categories over the course of training, across all conditions this learning trajectory was not modulated by the presence or lack of label as feedback on individual trials. Similarly, transfer-learning trajectories were not significantly affected for those participants whose initial training included verbal labels, for better or for worse. The inability of the current data to replicate previous findings on the influence of verbal labels in category learning draws into question the extent to which the Label Feedback Hypothesis can be extended into categorization.

Previous studies that have demonstrated a positive influence of verbal labels have focused mostly on shape-based categories, including the study upon which the present study is based (e.g. Lupyan et al., 2007; Lupyan & Thompson-Schill, 2011). Very early in language learning, English-speaking children develop a bias towards categorizing labeled object categories based on shape (Yoshida & Smith, 2005; Colunga & Smith, 2005). It’s possible that, as shape-based categories are based on dimensions that are historically predictive for English language speakers, the effect of labels during this type of categorization would be stronger than for other types of learning. This is supported by findings from Brojde, Proter, and Colunga (2011), who demonstrated that verbal labels hinder category learning defined by texture or brightness. They argue that the advantage of label comes about only when the relevant dimension aligns with the relevant dimensions in previous similar tasks, which in the case of our English-speaking participants would be shape over features such as orientation and frequency of line.

Conclusion

The purpose of the current investigation was to assess the effect of verbal labels on the ability to flexibly adjust categorization strategies when faced with changes in the environment. Previous literature (Lupyan et al., 2007; Lupyan & Thompson-Schill, 2011) has demonstrated that verbal labels influence category learning, improving both speed of learning and strength of representation. Some have argued that this effect of verbal labels is a demonstration of the top-down modulation of labels during learning and therefore shows that verbal labels are directly involved in the learning of concepts and categories (Lupyan, 2009). In this theory, labels work as material symbols upon which categories are attached (Clark, 1996), and so take part in the category learning process, possibly by modulating selective attention (Goldstone, 1998; Goldstone & Steyvers, 2001). Others, however, have argued that verbal labels are simply a more maximal form of feedback, and are therefore simple a form a facilitation, separate from the categories themselves (Maddox et al., 2008). In order to tease these two views apart, the present study considered the role of verbal labels
in flexible cognition, more specifically the ability of individuals to flexibly adjust their categorization strategies.

Despite a replication of the effect of selective attention across stimuli dimensions, the previous finding of the positive effect of labels as feedback for category learning was not replicated. The failure to replicate a positive effect of label on category learning raises questions as to the generalizability of the label feedback hypothesis. Given the issues raised in the current study above, it appears that not all types of category learning benefit from the presence of verbal labels (see also Brojde et al., 2011). Similarly, there is no evidence that labels modulate selective attention in a way that would either help or hinder flexibly adjusting one's categorization strategies. There was however, an effect in a single transfer condition that demonstrates that labels may aid in recovery from transfer when the type of transfer does not involve a change in selective attention. In the 180 degree transfer condition, while labels did not have a positive effect on learning during training, labels did interact with accuracy immediately after transfer, allowing those who learned with labels to recover faster. Future endeavors could continue to develop an understanding of the relationship between concepts, categories, and the words we use to invoke them.

Acknowledgments
We would like to thank Bhuvana Narasimhan, Claire Sims for their comments on an earlier draft, and the members of the CU language project for assistance in collecting data.

References