LABELS FACILITATE LEARNING
OF NOVEL CATEGORIES

GARY LUPYAN

Carnegie Mellon University Department of Psychology,
Center for the Neural Basis of Cognition
342C Baker Hall,

Pittsburgh PA , 15217, USA

A major feature that sets language apart from other communication systems is the use of categorylabels—words. In addition to providing a means of communication, there is growing evidence that category labels play a role in the formation and shaping of concepts. If verbal labels help humans acquire or use category information, one can ask whether it is easier to learn labeled categories compared to unlabeled ones. Normal English-speaking adults participated in a category-learning task in which categories were labeled or unlabeled. The presence of labels facilitated the learning of unfamiliar categories and resulted in more robust category representations. The advantage for acquiring named categories was observed even though the category labels did not convey any additional information and all participants had equivalent experience categorizing the stimuli. This work provides empirical support for the idea of labels as conceptual anchor points (Clark, 1997).

1. Introduction

Humans are the only animals to associate conceptual categories with words. Clearly, this association allows linguistic communication. Additionally, words have been argued to stabilize abstract ideas in working memory (Clark, 1997) and make them available for inspection (Clark & Karmiloff-Smith, 1993; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Vygotsky, 1962). These hypotheses are seductive, in part, because they suggest that words may have evolved, in part, as a cognitive aid rather than just as a means of communication. At present, no empirical evidence exists showing that learning names for categories facilitates their acquisition and representation. The present work provides just such evidence.

The question of the effect of labels on category-learning is not new and has been investigated extensively in the context of children's language acquisition. For instance, Waxman and Markow have (1995) have shown that superordinate labels such as "vehicle" can serve as "invitations" for children to form the appropriate category. One may rightfully point out that it is through words that we come to know what categories are relevant and learning words can guide our attention to the meaningful aspects of our environment. The present question, then, is not whether labels facilitate category-formation because they point out

the relevant categories (i.e., the label "chair" suggests that chairs are a useful and relevant category of objects), but whether named categories are easier to acquire *because* they have a name. Given a set of new objects organized, and equal experience categorizing them, are people who learn names for the categories better able to categorize the objects?

In support of Clark's notion that labels may act to stabilize abstract ideas, recent cross-cultural findings have provided evidence that language is closely linked to the human ability to appreciate exact numerosities (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004) such that entertaining the concept "exactly 15" may rely on having a discrete label "15" (or the ability to create an exact label by iterative means). There is also evidence that in anomia and aphasia more generally—neuropsychological conditions affecting language and the ability to produce names—the ability to form and act based on category knowledge may be severely compromised (De Renzi, Faglioni, Savoiardo, & Vignolo, 1966; Hjelmquist, 1989; Roberson, Davidoff, & Braisby, 1999). While such evidence is suggestive of cognitive roles played by words, a controlled comparison of category-learning with and without labels in normal English-speaking adults is currently lacking. If labels enable humans to form conceptual representations that would otherwise be more difficult or impossible to form by serving as "conceptual anchors," (Clark, 1997) one can ask whether it is easier to learn labeled categories compared to unlabeled ones. Crucially, because the present work relies on participants learning novel categories, it is possible to control stimulus familiarity (something not possible in the types of cross-cultural arithmetic investigations cited above) such that all participants have equal experience categorizing them, but some also learn names for the categories, while others do not.

2. Method

The task required participants to learn to classify 16 "aliens" (Fig. 1) into two categories—those to be approached, and those to be avoided—by responding with the appropriate direction of motion. These behavioral categories were chosen specifically because they exemplify typical categories learned by non-human animals. Participants were randomly assigned to a *label* or *no-label* condition. After each response, auditory feedback consisting of a buzz or a bell indicated whether the response was correct. In addition to the feedback, participants in the *label* condition saw one of two nonsense words ("grecious" or "leebish" depending on the alien category) appear next to the alien. The word appeared only after a response was made and was unrelated to the response.
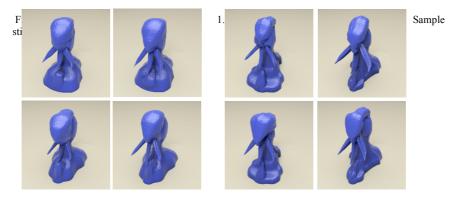
Participants in the *label* group learned that aliens that look a certain way should be avoided, *and* that they are *leebish*, while participants in the *no-label* group learned only that aliens looking a certain way should be avoided.

2.1. *Participants*

Forty-four Carnegie Mellon University undergraduates participated in the experiment for course credit. Data from 4 participants were excluded for failure to follow instructions. Data for the testing part of the task were available only for the last 20 participants.

2.2. *Materials*

The stimuli were a subset of the YUFO stimulus set (Gauthier, James, Curby, & Tarr, 2003). See Figure 1 for sample stimuli. The category distinction involved subtle differences in the configuration of the "head" and "body" of the creatures. The stimuli were presented on a 17" CRT screen. Responses were collected using a Gravis gamepad controller.

Figure 1. Sample stimuli

2.3. *Training*

Participants were told that they were acting as an explorer on another planet learning about alien lifeforms. Their task was to figure out which aliens they should approach and which they should move away from. On each training trial, one of the 16 aliens appeared in the center of the screen. After a 500 msec delay, an outline of a character in a spacesuit (the "explorer") appeared in one of four

quadrants—left, right, top, or bottom of the alien. Participants were instructed to respond with the appropriate direction key depending on the category of the alien. For instance, if the explorer appeared to the top of the alien, and the participant thought the correct response was to move towards, she would press the 'down' key, which would move the explorer down—towards the alien. Auditory feedback—a buzz or a bell—sounded 200 ms after each response. In the *label* condition, a printed label then appeared next to the alien. After another 1500 ms, the stimulus was erased and a fixation cross marked the start of the next trial. The total duration of each trial was equal in the two conditions. The pairing of the label ("leebish" or "grecious") with the category (move away, move towards), and the perceptual stimuli (family 1, family 2), was counterbalanced across participants. There were 144 training trials. Note that all participants received the same number of categorization trials, the only difference being in the presence or absence of category labels accompanying each response.

2.4. *Testing*

After completing the training trials, participants were instructed that in the upcoming part, no feedback (or labels) would be provided. On each trial, one of the aliens appeared in the center of the screen and remained visible until a response was made. Participants were instructed to press one button if they thought the alien belonged to the *escape* category and another button for *approach*. There was no mention of the names learned in the *label* condition. To determine whether participants were responding based on category knowledge or memory for specific examples, the testing part included novel stimuli from the learned categories in addition to previously categorized stimuli.

3. Results

A full treatment of the results will be reported elsewhere. All tests reported as significant were significant at the $p < .05$ level. Participants in the label condition performed significantly better overall and learned the categories significantly faster than participants in the no-label condition. After completing the learning phase, participants' knowledge of the categories was tested without feedback or labels. Since no reinforcement or correction was provided in the testing phase, category knowledge may deteriorate over time. It was found that participants in both groups did well above chance on the generalization trials, so the reported data collapses across both types of testing stimuli, a total of 144 trials. Participants who learned the categories with labels retained their category

knowledge throughout the testing phase, whereas those who learned the categories without labels displayed decreasing accuracy over time as revealed by a significant block by condition interaction (Figure 1B). There was no difference in latencies between the two groups in any part of the experiment, suggesting that the advantage in the label condition did not arise from greater time spent on task.
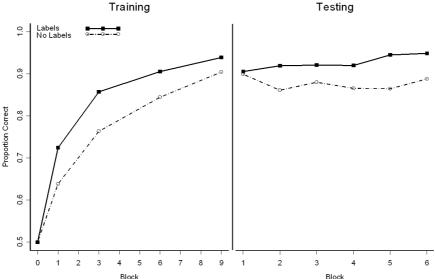

Fig 2. Mean accuracy for learning trials (left) and testing trials with no feedback or labels (right). Block 0 is included for illustration only and is not included in the analyses.

4. Discussion

Participants who learned names for novel categories showed faster learning of the categories compared to participants who categorized the items without learning names. Once the categories were learned by both groups, participants in the *label* group showed more robust category knowledge as evidenced by continued good performance when feedback and labels were removed.

The mounting evidence of effects of language on thought has been argued to be an artifact of cultural differences that limit exposure to certain types of cognitive tasks rather than arising from linguistic experience (Gelman & Butterworth, 2005; Li & Gleitman, 2002). The present task contrasted adults from similar backgrounds and controlled the exposure and categorization practice with the experimental stimuli. Despite undergoing the same amount of experience categorizing perceptual stimuli into behavioral categories (move

away or move towards), the presence of labels helped adults to represent the stimuli in terms of the appropriate behavioral categories.

Humans are unique in their ability to habitually associate words with mental categories. Most notably, this ability allows for linguistic communication. The present results suggest that learning category labels also provides for facilitated representation of the labeled concepts. Words or other discrete modes of representation may be necessary for entertaining certain abstract concepts like large exact numerosities. However, learning to associate words even with more ordinary perceptually-based categories such as those used here, facilitates their acquisition and results in more robust subsequent knowledge.

The categories used in the current experiment were by no means incommensurate with linguistic coding. Indeed, many participants in the *no-label* condition reported self-generating labels during the learning task. It remains to be seen whether labeling can produce an even greater difference in learning unfamiliar categories for which the self-generation of labels is more difficult or impossible.

4.1. *Why did the labels help?*

The current results do not provide an unambiguous explanation of the mechanism involved. However, we can rule out several explanations. Labels did not encourage participants to work harder, at least as evidenced by a lack of differences in the reaction times of the two groups. Alternatively, it might be argued that the presence of the written labels directed the participants' attention to a certain part of the stimulus which made it easier to find the features relevant to the categorization task. It was found, however, that performance was identical whether the labels were presented in the visual (written) or purely auditory modality. We have preliminary evidence showing that mere exposure to the labels does not provide the benefit. They need to be learned. The minority of participants who did not learn the mapping between the labels and the stimuli, or indicated on a post-study questionnaire that the labels were not helpful were comparable in their performance to those in the *no-label* condition.

What was it about the labels then that facilitated category-learning? The current task required participants to learn a subtle and difficult-to-verbalize distinction based on experience with individual category exemplars. In this context, the labels can be thought to provide perceptually-simple correlates to an otherwise perceptually-complex task. The labels seem to allow participants to represent the category distinction in terms of the categorical distinction:

"leebish" vs. "grecious," rather than the fuzzier perceptual distinction: "slightly more rounded and larger" vs. "less rounded and smaller."

4.2. *Are words special?*

Do labels need to be words or can any discrete information correlated with categories result in facilitated learning? A follow-up experiment was conducted that substituted word-labels with hieroglyphic symbols. Surprisingly, it was found that once the mapping between the symbols and categories was learned, they appeared to facilitate category-learning nearly as much as the verbal labels. It is not clear, however, whether the participants thought of the symbols as verbal labels. On the other hand, not all correlated information seems to have the facilitating effect. In another study, participants were told that the two groups of aliens lived in different places, and after making a response the alien would move towards one of these places (up or down). Learning to associate the motions with the aliens did not appear to facilitate category-learning. On the contrary, there is evidence that being exposed to this extra correlation actually made the task more difficult.

To summarize, we do not believe that only words can serve as discrete category markers. The labeling advantage may extend to other discrete environmental cues that are highly correlated with the categories. Words are special in that they normally constitute these discrete environmental cues, and so participants may find it more natural to use verbal labels as cues to categories compared to other types of information (e.g., lives above vs. lives below).

5. Conclusions

Providing redundant labels facilitated learning of novel categories and resulted in more stable category representations. Despite undergoing the same amount of experience categorizing perceptual stimuli into behavioral categories (*move away* or *move towards*), the presence of labels (nonsense words: "leebish"/"grecious") helped normal English-speaking adults to represent the stimuli in terms of the appropriate behavioral categories and resulted in more robust category representations. The current work shows that learning to name items facilitates placing them into appropriate behavioral categories, supporting Clark's (1997; 1998) view of words as "conceptual anchors" that facilitate negotiating perceptually-complex environments.

References

Clark, A. (1997). *Being There: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P.Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary themes* (pp. 162-183). Cambridge University Press.

Clark, A. & Karmiloff-Smith, A. (1993). The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought. *Mind & Language, 8,* 487-519.

De Renzi, E., Faglioni, P., Savoiardo, M., & Vignolo, L. A. (1966). The influence of aphasia and of the hemispheric side of the cerebral lesion on abstract thinking. *Cortex, 2,* 399-420.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20,* 507-523.

Gelman, R. & Butterworth, B. (2005). Number and language: how are they related? *Trends in Cognitive Sciences, 9,* 6-10.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306,* 496-499.

Hjelmquist, E. K. E. (1989). Concept-Formation in Non-Verbal Categorization Tasks in Brain-Damaged Patients with and Without Aphasia. *Scandinavian Journal of Psychology, 30,* 243-254.

Li, P. & Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition, 83,* 265-294.

Pica, P., Lemer, C., Izard, W., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306,* 499-503.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71,* 1-42.

Rumelhart, D. E., Smolensky, D. E., McClelland, J. L., & Hinton, G. E. (1986). Parallel Distributed Processing Models of Schemata and Sequential Thought Processes. In J.L.McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing Vol II* (pp. 7-57). Cambridge, MA: MIT Press.

Vygotsky, L. S. (1962). *Thought and Language.* Cambridge, MA: MIT Press.

Waxman, S. R. & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 257-302.