NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

DEPARTMENT OF PHILOSOPHY AND HISTORY OF SCIENCE

Assessing the interplay between language and category learning: the effects of verbal labels for categorization items and for the formed categories.

by

Fotis A. Fotiadis

Dissertation Thesis

Supervisory Committee

Athanassios Protopapas

Martijn Meeter

Nikolaos Smyrnis

Athens, Greece

May 2017

The thesis is approved

Athanassios Protopapas

Associate Professor

Department of Philosophy and History of Science, National and Kapodistrian University of Athens

____

Martijn Meeter

Professor

Department of Educational and Family Studies, VU University Amsterdam

____

Nikolaos Smyrnis

Associate Professor

Medical School, National and Kapodistrian University of Athens

____

This page was intentionally left blank.

Publications and Presentations

Peer-reviewed journal publications

Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory and Cognition 42 (1)*: 112–125.

Publications in conference proceedings - International

Fotiadis, F. A., & Protopapas, A. (2016). No effect of verbal labels for the shapes on Type II categorization tasks. 38th Annual Meeting of the Cognitive Science Society. Philadelphia, Pennsylvania, August 10–13.

Fotiadis, F. A., Protopapas, A., & Vatakis, A. (2011). The effect of cue naming in probabilistic category learning. In B. Kokinov, A. Karmiloff-Smith, & N. J. Nersessian (Eds.), European Perspectives on Cognitive Science: Proceedings of the European Conference on Cognitive Science – EuroCogSci 2011. Sofia, Bulgaria: New Bulgarian University Press.

Conference Presentations–Greek

Fotiadis, F, A., & Protopapas, A. (2016). [The effect of verbal labels for the categories on perception.] 4th Panhellenic Conference in Cognitive Psychology, Athens, Greece, April 15–17.

Fotiadis, F, A., & Protopapas, A. (2013). [The effect of verbal and nonverbal labels for the cues in probabilistic category learning.] 4th Panhellenic Conference in Cognitive Science, Athens, Greece, June 6–8.

Fotiadis, F, A., & Protopapas, A. (2013). [The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning.] 14th National Conference of Psychological Research, Alexandroupoli, Greece, May 15–19.

Fotiadis, F. A., Protopapas, A., & Vatakis, A. (2011). [The effect of visual cue nameability on a probabilistic category learning task.] 2nd Panhellenic Conference in Cognitive Science, Paros, Greece, June 2–5.

Fotiadis, F. A., Protopapas, A., & Vatakis, A. (2011). [The effect of visual cue nameability on a probabilistic category learning task.] 13th National Conference of Psychological Research, Athens, Greece, May 25–29.

Fotiadis, F. A., Protopapas, A., & Vatakis, A. (2010). [The effect of cue naming in probabilistic category learning.] 1st Panhellenic Conference in Cognitive Science, Athens, Greece, June 2–6.

This page was intentionally left blank.

Table of Contents

Chapter 1.	**Introduction**	7
Chapter 2.	**The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning**	15
Chapter 3.	**No effect of verbal labels for the shapes on Type II categorization tasks**	53
Chapter 4.	**Category labels facilitate learning and affect subsequent processing of category-diagnostic perceptual features**	70
Chapter 5.	**General discussion**	115
Appendix.	**Summary**	129
	Acknowledgements	131

1

Introduction

Category learning is a faculty instrumental for survival that enables human and nonhuman animals "dissect" and thus comprehend the world. The ability to classify stimuli contributes to the development of a broad range of capacities, from language learning (e.g., Lively, Logan, & Pisoni, 1993) to visual object understanding (e.g., Palmeri & Gauthier, 2004). Learning to categorize also supports the functioning and development of higher-level cognition, as it has been argued to form the basis of abstract thought (Goldstone & Hendrickson, 2010; Sloutsky, 2010).

The study of categorization is therefore an endeavor that may help elucidate the principles and cognitive processes of learning that are instrumental in cognitive development. To be more specific, the studies presented here suggest experimental manipulations capable of boosting learning. Indeed, to foreshadow the main conclusions of the present dissertation, an important finding is that verbal labels both for the categorization items and also for the formed categories were found to facilitate learning. Although these results were obtained in laboratory settings with adult participants, they could form the basis for educational research programs incorporating interventions. Thus, the research presented here may be said to be "basic," but the reported findings arguably have applications in education.

In the present dissertation I sought to investigate the interplay between the language faculty and category learning. Notwithstanding the fact that human and non-human categorization behavior is in some cases comparable (Smith, Minda, & Washborn, 2004), human categorizers may be strongly influenced by the language faculty (Lupyan, Rakison, & McClelland, 2007). I investigated the notion that language penetrates learning processes and modulates the learning of categories, with a specific focus on verbal labels. To this end, I utilized the methodology of studying the learning of artificial categories (Ashby & Maddox, 2005) and experimentally explored what verbal labels do when present at the two ends of the categorization spectrum. In particular, I explored (a) the effect of verbal labels for the features of categorization items, and (b) the effect of verbal labels for the formed categories.

Verbal Labels for the Stimuli

The first line of research was inspired by dual-systems theories of category learning. According to these theories the learning of categories is mediated by two distinct systems. A verbal, declarative, or explicit system is thought to underlie the learning of rule-defined categories, whereas a nonverbal, procedural, or implicit system is thought to be engaged in the learning of information-integration (or similarity-based) categories (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Minda & Miles, 2010). Computational models implementing the assumption of distinct systems have been successful in accounting for human behavior (e.g., the COVIS model, Ashby et al., 1998, or the ATRIUM model, Erickson & Krushke, 1998, but see e.g., Newell, Dunn, & Kalish, 2011, for opposing accounts). One of the key assumptions of dual-systems theories is that the two purported systems operate in parallel, and that the most successful system (in terms of correct classification decisions) assumes response delivery (e.g., Ashby et al., 1998, Shohamy, Myers, Kalanithi, & Gluck, 2008). My research question did not examine any aspect of the interaction or competition between distinct systems of category learning, it was rather inspired by the functional characteristics of the declarative system of categorization and the corresponding categorical structures.

Based on the description of rule-defined categories as being characterized by verbal rules of category membership (e.g., Gluck, Shohamy, & Myers, 2002; Minda & Miles, 2010), it was reasoned that verbal labels for the categorization items, or for the items' diagnostic features should boost explicit processes of rule discovery. In simpler words, it was assumed that participants would benefit in discovering and applying verbal rules of category membership when categorization stimuli are readily nameable compared to when the to-be-categorized material is hard-to-name.

The idea of rule-discovery facilitation due to names was initially examined using an auditory version of the weather prediction task (Knowlton et al., 1994), as detailed in Chapter 2. Learning in this task has been argued to be mediated by explicit processes of rule discovery, at least for young healthy participants (Shohamy et al., 2008). Moreover, boosting explicit processes has been shown to be accompanied by higher categorization accuracy (Price, 2009). To manipulate the availability of names for the stimuli I used hard-to-name auditory tones and trained separate groups of participants for three consecutive days to associate the cues to pseudowords or to hard-to-name ideograms. A third group was trained to associate stimulus intensity to color, and a fourth group remained unexposed to the cues. On the fourth day all participants were administered the same version of the weather prediction task utilizing the trained toned as cues. Results revealed group-level differences in post-training categorization accuracy, and—critically—higher categorization accuracy of the label compared to the ideogram group, suggesting that names for the cues facilitate explicit processes of category learning. These results are—to the best of my knowledge—the first evidence suggesting that human categorizers might benefit from names for the to-be-categorized material.

I sought to further test the idea of rule-discovery facilitation due to names, as detailed in Chapter 3. The manipulation of nameability in this experiment concerned the categorized items' features, using an experimental paradigm in the visual rather than the auditory modality, and with a deterministic rather than probabilistic category structure. In particular, I used the Type II category structure (Shepard et al., 1961) which is characterized by a verbal rule of category membership (Minda, Desroches, & Church, 2008). Facilitating explicit processes has been shown to lead to higher accuracy in the task (Minda et al., 2008). Categorization items in this experiment were composed of hard-to-name shapes (Vanderplas & Garvin, 1959) and similarly to Chapter 1 the availability of names for the values of the shape dimension was manipulated through training. Separate groups of participants were trained for two consecutive days to associate the shapes to pseudowords or to hard-to-name ideograms, whereas a control group received mock training and remained unexposed to the shapes. Results, contrary to the predictions, revealed no group-level differences in categorization accuracy.

In contrast to the results presented in Chapter 2, the assumption of a verbal facilitation in explicit processes of category learning was not supported in Chapter 3.The discrepancy of results across Chapters 2 and 3 (taking into account some relevant results from Chapter 4) is taken up in more detail in General Discussion.

Verbal Labels for the Formed Categories

The second line of research of the present dissertation investigated the effect of verbal labels for the formed categories, and it was inspired by the label-feedback hypothesis (Lupyan 2012a; 2012b). Lupyan's theory postulated a labels-dependent mechanism of selective activation of category-diagnostic perceptual features. This mechanism has been argued to offer a label advantage during learning and also have long-term effects on attention.

In Chapter 4 I examined the purported label advantage during learning to categorize, by using a within-subjects experimental design and by manipulating linguistic activity through the graded nameability of the categories' labels. Additionally, I examined long-term effects of learning to categorize under verbal labels on attention (Tolins & Colunga, 2015). Based on previous reports of stable perceptual learning (e.g., Goldstone, 1994), learned attention following category learning (Goldstone & Styevers, 2001; Krushke, 1996), or the effect of learned labels for categories (Lupyan, 2012a), it was reasoned that stimuli that had previously been predictive of label categories would capture attention to a greater extent compared to stimuli that had previously been predictive of hard-to-name categories. To examine the sustained effects of category labels on attention mechanisms, the category-diagnostic perceptual features were used in posttraining test tasks, specifically in Type II categorization tasks (Experiment 2), and in a visual discrimination task, using eye tracking (Experiment 3). Finally, to test if the effects of labels—both initial and sustained—are specific to categorization, the learning of categories was contrasted with the learning of associations. Control groups of participants in Experiments 2 and 3 learned named and hard-to-name associations instead of categories. With respect to an initial effect of labels, it was reasoned that a label advantage during learning to categorize should be above any advantage during learning to associate. With respect to a sustained effect of labels on attention it was reasoned that a label effect should be specific to category training. It was therefore assumed that posttraining processing of stimuli that had previously been paired with labels would be comparable to the processing of stimuli that had previously been paired with hard-to-name symbols. Results showed that named categories were consistently learned more accurately than hard-to-name categories, replicating the label advantage during learning to categorize (Lupyan et al., 2007). Moreover, the label advantage during learning to categorize was found to be greater than during learning to associate, suggesting that the categorization-specific mechanism of selective activation might complement a general advantage due to the processing of verbal stimuli. With respect to the sustained effects of category labels on attention, in both Experiment 2 and 3 results suggested that the shapes that had previously been diagnostic of named categories captured attention to a greater extent compared to shapes that had been diagnostic of hard-to-name categories. These sustained effects of labels on attention were found to be categorization-specific, in that following associative learning there was no evidence of differential processing of diagnostic features of named compared to hard-to-name associations. Collectively, the results of three experiments described in Chapter 4 seem to provide support to the label-feedback hypothesis (Lupyan, 2012a; 2012b), and the assumption of the warping of perceptual space due to categorization (Goldstone, 1994).

The present studies are grounded on two different subfields of category learning research: the dual-systems theories (e.g., Ashby et al., 1998) and the examination of the interplay between perceptual and learning processes (Goldstone, 1994; Lupyan, 2012a). By combining these two research traditions, I was arguably able to assess the effect of language on category learning processes more broadly. The current research may be said to contribute to the more general question of the interplay between language and presumably non-linguistic cognitive faculties, such as learning and perception (for more on the language and thought debate, see Gleitman & Papafragou, 2013).

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*(2). 107–140. doi:10.1037/0096-3445.127.2.107

Gleitman, L. R., & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *Handbook of Cognitive Psychology* (pp. 504–523). New York, NY: Oxford University Press.

Gluck, M. A., Shohamy, D., & Myers, C. E. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic classification learning. *Learning & Memory, 9*, 408–418. doi:10.1101/lm.45202

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178–200. doi:10.1037/0096-3445.123.2.178

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69–78. doi:10.1002/wcs.26

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*(1), 116–139. doi.org/10.1037/0096-3445.130.1.116

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106–120. doi:10.1101/lm.1.2.106

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. Connection *Science, 8*(2), 225–248. doi:10.1080/095400996116893

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*(3), 1242–1255. doi:10.1121/1.408177

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition, 3*, 54. doi:10.3389/fpsyg.2012.00054

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Waltham, MA: Academic Press, doi:10.1016/B978-0-12-394293-7.00007-8

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18*, 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 52, pp. 117–162). San Diego: Academic Press. doi:10.1016/S0079-7421(10)52003-6

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed), *The Psychology of Learning & Motivation, 54*, 167–215. San Diego: Academic Press. doi:10.1016/B978-0-12-385527-5.00006-1

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*(4), 291–303. doi:10.1038/nrn1364

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42. doi:10.1037/h0093825

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews, 32*, 219–236. doi:10.1016/j.neubiorev.2007.07.008

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science, 34*(7), 1244–1286. doi:10.1111/j.1551-6709.2010.01129.x

Tolins, J., & Colunga, E. (2015). How words anchor categorization: conceptual flexibility with labeled and unlabeled categories‡. *Language and Cognition, 7*(2), 219–238. doi:10.1017/langcog.2014.26

Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology, 57*(3), 147–154. doi:10.1037/h0048723

2

The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning

This chapter has been published as:
Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory and Cognition 42(1)*: 112–125

Abstract

Learning in a well-established paradigm of probabilistic category learning, the *weather prediction task*, has been assumed to be mediated by a variety of strategies reflecting explicit learning processes, such as hypothesis testing, when administered to young healthy participants. Higher categorization accuracy has been observed in the task when explicit processes are facilitated. We hypothesized that furnishing verbal labels for the cues would boost the formation, testing, and application of verbal rules, leading to higher categorization accuracy. We manipulated the availability of cue names by training separate groups of participants for three consecutive days to associate hard-to-name artificial auditory cues to pseudowords or hard-to-name ideograms; or to associate stimulus intensity to colors; a fourth group remained unexposed to the cues. Verbal labels, cue individuation, and exposure to the stimulus set each had an additive effect on categorization performance in a subsequent 200-trial session of the weather prediction task using these auditory cues. This study suggests that cue nameability, when controlled for cue individuation and cue familiarity, has an effect on hypothesis testing processes underlying category learning.

Introduction

Categorization is a fundamental aspect of cognition underlying a broad range of human behaviors and skills such as language acquisition, inference, concept formation, and decision making. The cognitive neuroscience of category learning has extensively tried to shed light on its mechanisms, representational contents, and neural substrates. Alternative approaches suggest that category learning is mediated either by qualitatively distinct systems (Ashby & Maddox, 2011; Poldrack & Foerde, 2008) or by a single learning mechanism (Newell, Dunn, & Kalish, 2011).

Explicit Hypotheses in Category Learning

Multiple systems theorists draw a distinction between a declarative, explicit, or verbal system or pathway and a procedural, implicit, or non-verbal system (Ashby & Maddox, 2005, 2011; Minda & Miles, 2010; Poldrack & Foerde, 2008; Squire, 2004). The explicit system is thought to be engaged when hypothesis testing processes—such as the formation, testing, and application of a verbalizable rule or strategy—can lead to successful performance and the knowledge acquired is accompanied by awareness. The implicit system underlies performance when no verbalizable rules exist or can be easily applied, in which case integration of information across multiple trials occurs or perceptual learning processes are recruited. Knowledge acquired by the implicit system is considered unavailable to conscious recollection. The two systems have been suggested to compete (Ashby, Alonso-Reese, Turken, & Waldron, 1998; Poldrack et al., 2001) or operate in parallel (Dickerson, Li, & Delgado, 2011; Minda & Miles, 2010; Shohamy, Myers, Kalanithi, & Gluck, 2008).

Single system theorists, on the other hand, have questioned the parsimony of multiple categorization systems (Newell et al., 2011) and the validity of methodologies (e.g., double dissociations) utilized in the past (Newell & Dunn, 2008; Newell, Dunn, & Kalish, 2010). Instead, they have suggested that human categorization is achieved through a single general learning mechanism (Newell, Lagnado, & Shanks, 2007) and is accompanied by high levels of awareness for the learned material (Lagnado, Newell, Kahan, & Shanks, 2006). The hypothesis of multiple memory systems or pathways remains a matter of current debate in the study of categorization (e.g., Ashby & Maddox, 2011; Newell et al., 2011).

Regardless of the existence and functional independence of discrete categorization systems, few would argue against the notion that category learning employs—in at least some task structures—hypothesis testing processes (Ashby & Maddox, 2005), inner rehearsal (Lupyan, Rakison, & McClelland, 2007), or verbalizable strategies (Gluck, Shohamy, & Myers, 2002). Executive functioning mechanisms have been argued to contribute to category learning by means of formulation, testing, and application of verbal rules of category membership (Price, 2009). In particular, human category learning has been argued to be influenced by verbal processes (Minda & Miles, 2010) since "humans have the potential benefit of [verbal] labels" (Lupyan et al., 2007, p. 1077).

Although language in general seems to play an important role in category learning, researchers have mainly manipulated the category structure (i.e., the availability of an easily verbalizable rule) to examine the effect of verbal processes on categorization (Ashby & Maddox, 2005; Miles & Minda, 2011). Recently, Lupyan (2006; Lupyan et al., 2007) studied the influence of category labels. He showed that verbal labels—as opposed to location cues —facilitated categorization of artificial stimuli when paired with category classes. However, not much attention has been drawn to the existence of labels for the items to be categorized. It stands to reason that if the stimuli are accompanied by verbal labels then hypothesis testing or inner rehearsal processes will be facilitated, because participants would find it easier to form, test, and apply rules such as "respond 'rain' whenever the triangle card is present" (Gluck, Shohamy, & Myers, 2002, p. 416). In contrast, in the case of non-nameable stimuli it would not be so easy to explicitly state and apply rules concerning them.

In the present study we sought to test this idea using hard-to-name cues in the context of a prototypical probabilistic category learning task. Participants were first trained to learn novel nonsense verbal labels or other hard-to-name pairings for the cues. They were subsequently administered the category learning task using these cues, in order to explore the effects of cue nameability on learning to categorize.

The weather prediction task

The prototypical weather prediction task (WPT; Knowlton, Squire, & Gluck, 1994) is a perceptual categorization task based on a paradigm developed by Gluck and Bower (1988).

Participants are asked to classify combinations (patterns) of four cards with geometric shapes (cues) into one of two possible outcomes, namely “sun” and “rain.” The task has a probabilistic structure in that each cue is associated with an outcome with a fixed probability. Two of the cues are highly predictive and the other two are less predictive of a specific outcome. Overall, throughout training a combination of cues may predict one outcome on some trials while on other trials the same combination may predict the alternative outcome (see Method). Corrective feedback is provided after every trial. It is now well established that both healthy and brain damaged participants gradually improve in categorization accuracy in a variety of versions (i.e., visual stimuli serving as cues, and category classes) of this task (e.g., Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004; Knowlton et al., 1994).

The WPT has been widely used by multiple systems theorists to assess the relative contribution of explicit (declarative) and implicit (procedural)[1] learning processes to the acquisition of knowledge (Poldrack & Rodriguez, 2004). Early neuropsychological studies suggested that the task mainly taps procedural learning processes (Knowlton, Mangels, & Squire, 1996; Knowlton et al., 1994; Reber, Knowlton, & Squire, 1996). However, neuroimaging studies (Poldrack et al., 2001), mathematical modeling of healthy participants' behavior (Gluck et al., 2002), and re-examination of clinical populations' behavior (Hopkins et al., 2004; Shohamy, Myers, Onlaor, & Gluck, 2004) have indicated an engagement of both declarative and procedural processes, presumably at different periods in training.

The mathematical modeling of young healthy participants' behavior has suggested that, early in the task, participants use sub-optimal verbalizable strategies (Gluck et al., 2002; Meeter, Myers, Shohamy, Hopkins, & Gluck, 2006; Meeter, Radics, Myers, Gluck, & Hopkins, 2008) that can be said to be declarative (Shohamy et al., 2008). Later in training participants shift to optimal multicue strategies. These later strategies have also been suggested to be accompanied by high levels of self-insight (Lagnado et al., 2006) or awareness (Price, 2009) and thus can be said to reflect explicit processes as well. Newell et al. (2007) suggested that the task is mediated by a single explicit learning mechanism.

1 The terms *declarative* and *procedural* have been used to denote memory systems (e.g., Squire, 2004) whereas the terms *explicit* and *implicit* learning denote processes assessed by direct or indirect experimental tests of knowledge (e.g., Reber & Johnson, 1994). Some researchers use *declarative* and *explicit,* as well as *procedural* and *implicit,* interchangeably (Price, 2009), in an effort to reconcile memory systems and learning processes approaches.

Similarly, Poldrack and Foerde (2008) suggested that normal young adults may use declarative learning strategies to solve the task. Thus, although the WPT is a legacy of the multiple systems field, recent research suggests that young healthy participants' behaviour is mediated by explicit learning processes entailing hypothesis testing of verbal rules (Price, 2009).

Researchers have experimentally manipulated the engagement of explicit processes during the WPT. Gluck et al. (2002) tested young healthy participants in two versions of the WPT. When the cue-outcome contingencies were less probabilistic (in their Experiment 2)—a manipulation thought to encourage declarative mediation (Foerde, Knolwton, & Poldrack, 2006)—performance measures increased throughout training, compared to a more probabilistic version (in their Experiment 1). Secondary task demands were introduced during WPT training to hamper explicit processes, resulting in impairment of WPT categorization performance throughout (Foerde, Poldrack, & Knowlton, 2007) or at the second half of training (Foerde et al., 2006; Newell et al., 2007), compared to single-task conditions. More recently, Price (2009; Experiment 2) reduced the time available for feedback processing, in order to impair explicit processes. Performance of participants in the long-feedback version was consistently greater than in the short-feedback version. Thus, empirical data suggest that experimental manipulations favoring explicit processes result in higher categorization accuracy. Consistent with this interpretation, reduction in WPT performance is also observed in special populations thought to be less efficient or impaired in their declarative encoding, and thus less able to form test and apply verbal rules, such as older healthy participants (Abu-Shaba, Myers, Shohamy, & Gluck, 2001) or hypoxic patients with MTL lesions (Hopkins et al., 2004), respectively.

Design and rationale of the present study

In the present study we employed a cue-response trial-and-error training paradigm modeled on the WPT. We used computer-generated auditory tones as cues because the majority of people do not possess pre-established labels for tones (Galizio & Baron, 1976). Prior to the WPT procedure, two groups of participants received extensive training to associate four novel auditory cues to pseudowords (label training condition) or hard-to-name ideograms (ideogram training condition). A third group of participants were exposed to the same stimuli over the same number of trials but learned to associate sound intensity to hard-to-name colors (intensity training condition), disregarding cue identity. A fourth group remained unexposed to the auditory cues (no-training condition). All groups were subsequently administered an auditory version of the WPT (Fotiadis, Protopapas, & Vatakis, 2011) utilizing these cues.

The main hypothesis and motivation underlying our study is as follows: if verbal labels facilitate the formation, testing, and application of verbalizable rule-based strategies, and if facilitating explicit learning processes is accompanied by higher categorization accuracy (Price, 2009), then the label training group should outperform the ideogram training group in the WPT. However, the availability of verbal labels is not the sole potential facilitator of category learning, as it presupposes both familiarization and individuation, which may be partially responsible for any observed learning benefits. Cue-response training requires the formation of individuated representations for the cues, potentially causing participants to develop perceptual anchors (Ahissar, 2007). Such individuated representations may help stabilize representations in working memory and facilitate executive functions such as hypothesis testing. If this is the case, then participants in the ideogram training condition ought to have an advantage in WPT categorization accuracy compared to the intensity training group in which cue identity was instructed to be unattended and varied orthogonally to the intensity task. Finally, mere exposure to the stimulus features has been shown to affect subsequent categorization performance (Folstein, Palmeri, & Gauthier, 2010). We thus predicted that participants in the intensity training group would outperform the no-training group.

Method

Participants

Eighty five undergraduate and graduate students (19 male, M_{age} = 25.8, SD = 4.05) of the Philosophy and History of Science Department, University of Athens, Greece, were randomly assigned to one of the three training conditions, receiving course credit for participation, or volunteered. Due to technical failures in collecting the training data or participants' errors in following instructions, 10 participants were excluded from analysis. Thus, there were data from 23 participants (7 male, M_{age} = 27.7, SD = 4.47) in the label training condition, 22 participants (6 male, M_{age} = 24.3, SD = 2.55) in the ideogram training condition, and 30 participants (5 male, M_{age} = 25.6, SD = 4.35) in the intensity training condition. In addition, twenty graduate students (2 male, M_{age} = 20.3, *SD* = 3.5) from the Psychology Department, Panteion University, Athens, Greece, were administered only the WPT (no-training group). All participants reported normal hearing and normal or corrected-to-normal vision, no history of neurological illness, and no dyslexia diagnosis.[2]

Materials

Cues. Four 300-ms long frequency-modulated tones, similar to those used by Holt and Lotto (2006), served as cues. The tones were created in Carnegie Mellon University using parameters listed in Table 1. A pilot study employing a 2AFC intensity discrimination task indicated that high-pitched tones were perceived as louder compared to low-pitched tones. Because of the need to be used in intensity training, the four tones were perceptually equated in intensity. Perceptual equation (outlined in the online supplement) resulted in the tones' *adjusted* intensity levels, subsequently used in the training procedure.

Four intensity levels were additionally created for each tone: the *highest* intensity corresponded to the tone's adjusted level, while the *high, low,* and *lowest* levels were created by decrements of 3, 6, and 9 dB down from the adjusted level, respectively. The 3 dB step was determined in pilot experiments aiming to equate—to the extent possible—training performance in the 3 conditions.

In the WPT the original (unadjusted) tones were used with all 4 groups.

Table 1

Carrier and Modulation Frequency of the Four Tones that Served as Cues

Tone	Carrier frequency (Hz)	Modulation frequency (Hz)
1	790	360
2	1060	360
3	790	198
4	1060	198

2 Dyslexia was a concern because it has been linked with impaired learning of audio-visual pairing (Hulme, Goetz, Gooch, Adams, & Snowling, 2007).

Pseudowords. Four Greek pseudowords were created to serve as new names for the tones, namely σάβης (/'savis/), λίμης (/'limis/), ρήτης (/'ritis/), and δόθης (/'ðoθis/). They were equal in number of letters, syllables, phonemes, stress position, and orthographic typicality (mean orthographic Levenshtein distance of the 20 nearest neighbors—OLD20—was 2.00 for all cues, taking stress into account, and between 2.15–2.85, ignoring stress; Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012; Yarkoni, Balota, & Yap, 2008).

Ideograms. Four Chinese characters were selected, based on (a) number of strokes and (b) structure (a single component; Yan, Qiu, Zhu, & Tong, 2010): 豸(U+8C78), 赤 (U+8D64), 辛 (U+8F9B), and 辰 (U+8FB0). To equate perceptual salience, the first character was rotated to the right by 20 degrees. A stroke was erased from the fourth character, resulting in 7 strokes for each of the final stimuli, shown in Fig. 1.A.

Colors. Three "hard-to-name" colors (RGB: 0x649EA7, 0x583232, 0xBFBC8F) were sampled from the on-line version of a study used to assess the involvement of language processing brain regions in a perceptual decision task (Tan et al., 2008; not implying that our stimuli were identical to those used in their study, due to lack of chromatic calibration). A fourth color (0xFEAD5C) was selected, subjectively judged to be hard-to-name. All color stimuli are shown in Fig. 1.B.

Procedure

Participants in the training conditions received instructions, a set of headphones, and a questionnaire (in the ideogram and intensity training conditions) on or before the first day of training. Moreover, each participants' computer volume was calibrated (see the online supplement for details). Training took place unsupervised at home for three consecutive days. Compliance was monitored daily by email or phone and by inspection of the data. On the fourth day, the WPT was administered at the university lab. Participants used headphones during the tasks.

Training. The training tasks and all following procedures were programmed in DMDX display software (Forster & Forster, 2003). Trial randomization was done with Mix (Van Casteren & Davis, 2006).

Verbal label training. There were 192 trials in each training session. Each cue was presented 12 times in each of 4 intensity levels. Participants heard one tone in each trial. They were instructed to guess at first, gradually learning the correct response for each tone through corrective feedback. They were explicitly told that the purpose of the task was to learn “a name for each tone” and not just make the correct response. On the first day of training they were asked to read aloud the word before responding. The correspondence between sounds and pseudowords was randomly selected for each participant.

Trial structure is shown in Fig. 1.C. A cross appeared on the center of the screen for 500 ms. A tone lasting 300 ms followed, simultaneous with the 4 response options (pseudowords) presented on screen in a vertical configuration. On the first day of training there was an additional latency period of 500 ms after the presentation of the tone, during which participants were to pronounce the word. Pseudowords remained on screen for up to 5 s, until a mouse click on one of them. Response feedback was provided for 500 ms (“correct,” “wrong,” or “no response”). The intertrial interval was 1 s.

Trial order was pseudorandom, fixed for all participants, but different for each day of training. Randomization constraints precluded (a) the same configuration of response cues on two consecutive trials, (b) a distance between trials with the same tone (regardless of intensity) less than 2, and (c) a distance between trials with the same intensity less than 1. There was a short break halfway through the procedure. Training lasted on average 18 minutes on the first day and 15 minutes on the second and third day. Training tasks were conducted online using DMDX remote testing mode.

Figure 1. Training procedure and stimuli. (A) Symbols used as response cues for the ideogram-training condition. (B) Colors used in the intensity-training condition. (C) Sequence of events in a training trial. Response 1, 2, 3 and 4 are used here to depict the four available response options and were replaced with pseudowords, stimuli depicted in (A), and stimuli depicted in (B) in the label, ideogram, and intensity training conditions respectively. The symbol ♪ was never presented.

Ideogram training. Ideogram training was identical to verbal label training except that (a) 4 ideograms (randomly paired with tones for each participant) replaced the 4 pseudowords, (b) participants were instructed to learn the ideogram that corresponded to each of the tones, and (c) there was no delay to pronounce the labels on the first training day. Participants were instructed to fill in the sealed questionnaire received at the initial meeting on completion of the third day's training. In this questionnaire the four ideograms were printed and participants were asked to name them using only one word.

Intensity training. Participants in intensity training heard the same stimuli as in the other training conditions but were asked to learn the color that matched each intensity level. They were explicitly instructed to ignore the identity of the tones and only pay attention to intensity. Intensity-color correspondence was randomized across participants. All other aspects of the procedure were the same as in the ideogram training condition. Following third day's training participants were asked to fill in a questionnaire asking for the names of the four colors using one word (as in Sturges & Whitfield, 1995).

WPT. Participants were told that they would take part in a learning experiment and would be asked questions about it at the end. They were not informed of the probabilistic nature of the task. For those in the training conditions it was noted that this was neither a continuation nor a test of their training. Written instructions were presented on the screen (adopted from Lagnado et al., 2006). Five practice trials were given before the actual experiment, for familiarization and sound volume adjustment, using animal sounds as cues.

The probabilistic structure of this auditory version of the WPT followed that of Gluck et al. (2002, Experiment 2). As already noted, each cue is independently associated with an outcome with a fixed probability. This probability can be calculated from Table 2 (as described by Shohamy et al., 2004). For example, Cue 1 is present in patterns H to N, which appear in 100 out of 200 trials of the experiment. In these 100 trials the outcome of sun occurs 20 times and the outcome of rain occurs 80 times. Thus, Cue 1 is associated with sun with probability 20 ÷ 100 = .2 and with rain with probability .8. Likewise, it can be calculated that cues 2, 3, and 4 predict sun with probabilities .4, .6, and .8 respectively. Cue 1 and Cue 2 are therefore predictive of sun, Cue 3 and Cue 4 are predictive of rain, and the highly predictive cues of the task are Cue 1 and Cue 4 for sun and rain respectively. The assignment of tone (Tone 1, Tone 2, etc.) to associative strength (Cue 1, Cue 2, etc.) was counterbalanced across participants, and the relative position of a tone within a pattern was held constant for a given pattern and a given participant.

Table 2

Pattern and Outcome Frequencies of the Weather Prediction Task

	Cue Present						
Pattern	1	2	3	4	Sun	Rain	Total
A	0	0	0	1	17	2	19
B	0	0	1	0	7	2	9
C	0	0	1	1	24	2	26
D	0	1	0	0	2	7	9
E	0	1	0	1	10	2	12
F	0	1	1	0	3	3	6
G	0	1	1	1	17	2	19
H	1	0	0	0	2	17	19
I	1	0	0	1	3	3	6
J	1	0	1	0	2	10	12
K	1	0	1	1	5	4	9
L	1	1	0	0	2	24	26
M	1	1	0	1	4	5	9
N	1	1	1	0	2	17	19
Total					100	100	200

Note. 1 = cue present, 0 = cue absent.

In each trial a series of tones forming a cue pattern were delivered through the headphones sequentially, with an intercue interval of 1 s. Hence, the duration of each pattern ranged from 0.3 s (1-cue pattern) to 2.9 s (3-cue pattern). Following an additional interval of 1 s, two icons representing the outcomes (a sun and a raining cloud) appeared on the screen for the participant to respond by pressing the corresponding key on the keyboard. At registration of a response, the correct outcome was presented on screen for 2 s along with feedback: a happy smiley and a high tone (frequency: 1000 Hz, duration: 0.1 s) for correct selection, or a frowning smiley and a low tone (frequency: 500 Hz, duration: 0.1 s) when incorrect. If the participant did not respond within 2 s, a “Please respond now” prompt appeared on the bottom of the screen. The trial was terminated if no response was registered within 5 s total, counting as “incorrect” for the purpose of analysis. Following Knowlton et al. (1994), a yellow bar on the right side of the screen provided a rough estimate of performance. The intertrial interval was 500 ms. Short breaks were given every 50 trials. The complete sequence of events in a 2-cue auditory pattern trial is shown in Fig. 2. The duration of the categorization task was 35 minutes on average.

Figure 2. Sequence of events in a 2-cue auditory pattern trial of the WPT, yielding the “rain” outcome, along with the two possible types of feedback. A “Please respond now!” prompt appeared on screen if the participant did not respond within 2 s of the presentation of the possible outcomes. The icon ♪ was never presented.

Cue naming. Immediately after the WPT, participants were asked to write down which single cue they considered most likely for each outcome (the precise formulation of the questions was based on Reber et al., 1996). Participants in the three training conditions were also presented with the four tones again, and were asked to denote which tone corresponded to their two previous responses.

Data Analysis.

Analyses reported below (except for cue naming) employed generalized mixed-effects logistic regression models for binomial distributions (Dixon, 2008) via a logit transformation (Jaeger, 2008), with participants and stimuli (or patterns of auditory stimuli for WPT) as random factors (Baayen, Davidson, & Bates, 2008), fitted with restricted maximum-likelihood estimation using package lme4 (Bates & Sarkar, 2007) in R (R Development Core Team, 2011). Effect sizes (β) are estimated log odds regression coefficients, with zero corresponding to no effect.

Training. Training data were analyzed in terms of correct or erroneous responses.

WPT. Following standard procedure, participants' categorization performance was measured in terms of optimal responding (Knowlton et al., 1994). A response was marked correct if it corresponded to the most likely outcome given task contingencies, regardless of the actual feedback presented to the participant on that particular trial. For example, throughout the task trials incorporating pattern A were marked as correct if and only if the response was "Sun." As can be seen in Table 2, patterns F and I are equally associated to both outcomes, hence no optimal response can be defined for them. Responses to these patterns (12 trials overall for each participant) were not included in the analysis.

Cue naming. Answers were scored with 1 if participants responded with the tone that was highly predictive of the stated outcome, with 0.75 for the less predictive tone, 0.50 and 0.25 for the tones predictive of the opposite outcome, weakly or strongly, respectively, and 0 for no answering. Cue selection performance was the sum of the two outcomes, ranging from 0 to 2.

Results

Training

Performance increased throughout and across the three days of training, but not all participants exhibited high performance at the end of the third day. To ensure that subsequent categorization performance (on the WPT) would be subject to the trained cue associations, we excluded participants exhibiting low performance (45% or less) in the second half of the third day of training. This included two "non-learners" in label training, two in ideogram, and seven in intensity. Moreover, to equate sample size across conditions, we randomly excluded one participant from the label condition and three from intensity (see Fig. S1 in the online supplement). Data shown and analyzed henceforth correspond to the following sample: 20 participants (6 male, M_{age} = 26.8, SD = 3.47) in the label training condition, 20 participants (6 male, M_{age} = 24.4, SD = 2.62) in ideogram training, and 20 participants (4 male, M_{age} = 25.7, SD = 3.92) in intensity training.

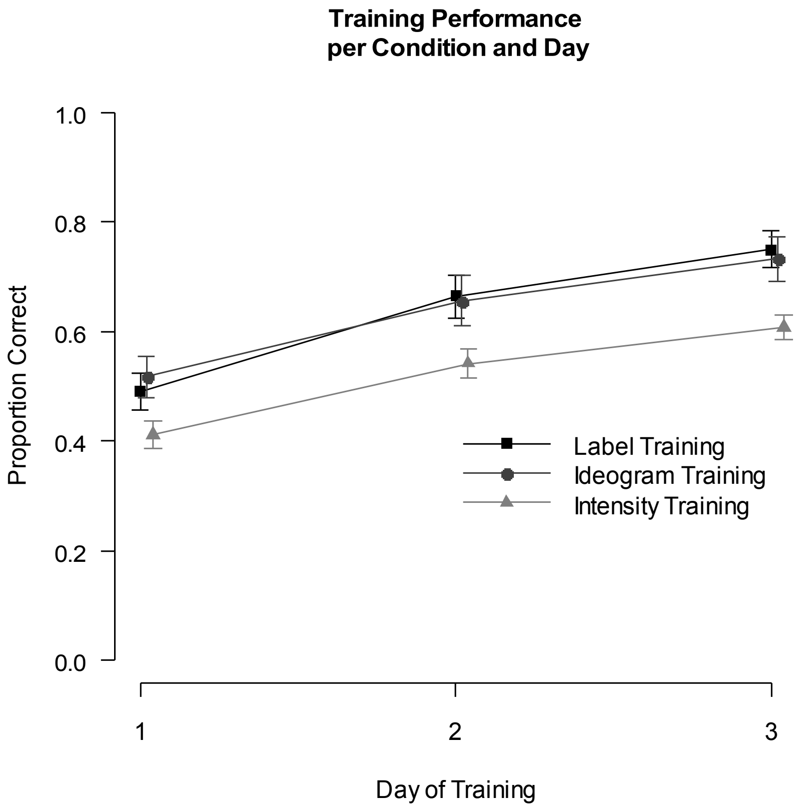


Figure 3. Mean accuracy of 60 participants (20 in each training condition) in cue-response training. Error bars show between-subjects standard error of the means.

Mean performance in training per condition and day is shown in Fig. 3. Participants' responses were analyzed with a model including fixed effects of trial, training condition, and day of training, as well as their interactions, and random effects of participants and of stimuli (four tones by four intensity levels, i.e., 16 distinct stimuli). In R notation, one such model was specified as:

```
accuracy ~ trial * condition * day + (1+trial|participant) +
(1|stimulus)
```

with two levels of accuracy ("correct" and "wrong") regressed onto 192 trials, three levels of condition ("intensity," "ideogram," and "label"), and three levels of day. By-participant random slopes of trial were included to model participants' individual learning rates; by-stimulus random slopes of trial did not improve model fit and were excluded. Quadratic effects of trial were not significant and were therefore excluded from the models.

The main purpose of the analysis was to assess whether training resulted in comparable knowledge—by the end of the third day of training—of the cue-response pairings across the three groups. Therefore, the model's intercept was set at the end of training (i.e., the levels of the day predictor were ordered as “day3” “day2” and “day1,” and trial was specified numerically as −191, −190, ..., −1, 0). The simple effect of condition indicated that the odds of correct responding at the end of Day 3 of training were comparable between the label and ideogram training conditions whereas both of these groups outperformed the intensity training group (label vs. ideogram: $\beta = -.180$, $z = -0.700$, $p = .484$; label vs. intensity: $\beta = .667$, $z = 2.633$, $p = 0.009$; ideogram vs. intensity: $\beta = .847$, $z = 3.332$, $p < .001$; the last two estimates survived Bonferroni correction for three pairwise comparisons). There was a marginal interaction of trial by condition, indicating that change in correct responding as trials progressed in Day 3 was marginally different between the label and ideogram conditions but comparable between the other conditions (label vs. intensity $\beta = -.001$, $z = -1.241$, $p = .215$; ideogram vs. intensity: $\beta = .001$, $z = 1.086$, $p = .278$; label vs. ideogram: $\beta = -.002$, I = −2.185, $p = .029$, not surviving Bonferroni correction for three comparisons). No three-way interaction survived Bonferroni correction for multiple comparisons.[3]

Written responses on the post-training questionnaire assessing ideograms' names confirmed that the symbols used were hard-to-name and did not invoke common associations. Names given were mainly idiosyncratic (such as “air” or “sunset”). A few (6 out of 20) participants named the ideograms after the sounds they had been paired to (i.e., they gave names such as “bass” or “shrill”).

In contrast, questionnaire responses regarding colors revealed participants' tendency to give common names to Color 1 (“light blue”—a single word in Greek—by 10 participants, “blue” by 7), Color 2 (“brown” by 10), Color 3 (“beige” by 8, “grey” by 7), and Color 4 (“orange” by 15).

WPT

Participants' performance is shown in Fig. 4 in blocks of 10 trials. Participants averaged 74.9% (*SD* = 8.7%) optimal responses over all 200 trials in the label training condition, 71.7%

3 Analysis of data at the end of Day 1 indicated increased accuracy of the ideogram training condition compared to the intensity condition, but comparable accuracy among the other conditions. Analysis of data at the end of Day 2 indicated higher accuracy of the label training condition compared to the intensity condition but comparable accuracy among the other conditions. All analyses are available from the authors upon request.

(*SD* = 8.8%) in the ideogram condition, 68.5% (*SD* = 12.0) in the intensity condition, and 63.6% (*SD* = 9.0) in the no-training condition.

Responses were analyzed with a model including fixed effects of target (optimal) response, trial, and training condition, as well as their interactions, and random effects of participants and of patterns of auditory cues. In R notation, the model was specified as:

```
response ~ target * trial * condition + (1+trial|participant)
+ (1|pattern)
```

with two types of response ("Sun" and "Rain") regressed onto two types of target ("Sun" and "Rain"), 188 trials (centered, thus specified numerically as, -99.5, -98.5,..., 98.5, 99.5, excluding trials presenting patterns "F" and "I"), and four types of condition ("no-training", "intensity", "ideogram", and "label"); there were also twelve types of pattern ("A"..."N", excluding patterns "F" and "I"). By-participant random slopes of trial were included to model participants' individual learning rates.

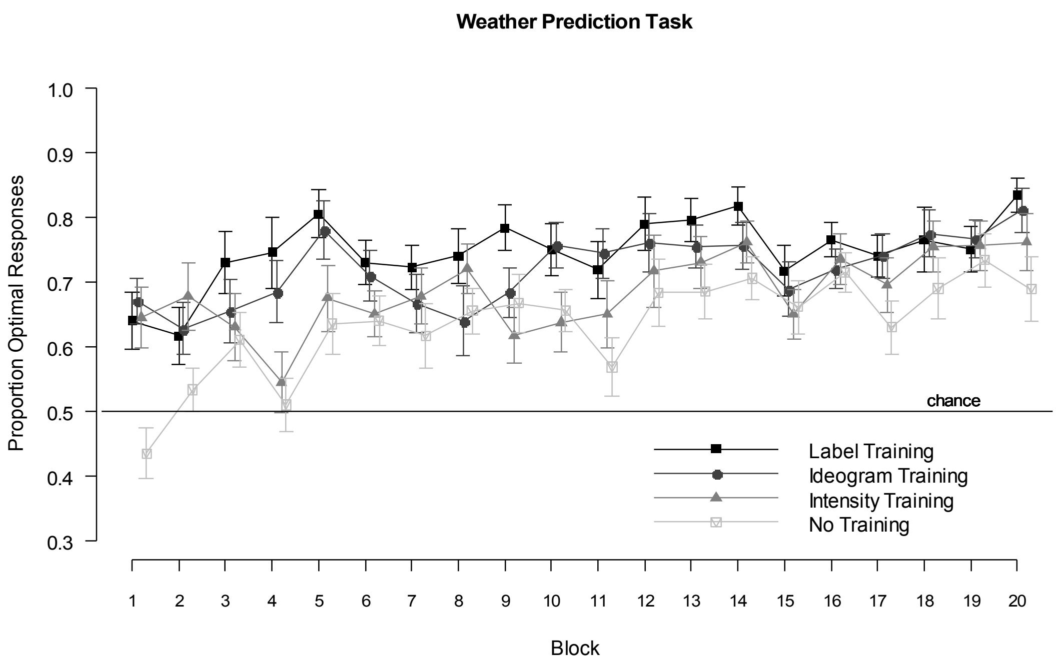


Figure 4. Post-training categorization performance of the four training conditions in blocks of 10 trials. The dotted line denotes chance performance (50%). Error bars show between-subjects standard error of the means.

In this model, learning effects would be evident as a significant interaction of trial by target, insofar as increases in trial would increase the probability of responding correctly. This interaction was significant (β = .010, z = 7.830, p < .001). A triple interaction including condition would indicate differential learning effects across training conditions, however this interaction was not significant for any pair of conditions (all β < .002, p > .3).

There were significant interactions of condition by target, indicating significant performance differences between conditions, in the following order: label > ideogram > intensity > no-training. Successive pairwise differences survived Bonferroni correction for three comparisons and were all highly significant (label vs. ideogram: β = .351, z = 3.205, p = .001; ideogram vs. intensity: β = .341, z = 3.247, p = .001; intensity vs. no-training: β = 0.451, z = 4.441, p < .001)[4].

Cue naming

In response to the post-categorization questionnaire most participants provided verbal descriptions of the tones related to their acoustical features, such as "the high-pitched one" or "the bass sound." In the label condition, 11 out of 20 participants used the trained pseudowords. In the ideogram condition, 4 participants gave descriptions related to the visual features of the ideograms, such as "the *F*" or "antenna." None of the participants in the intensity training condition used a color name to describe the tones.

Mean cue selection scores were 1.79 (SD = 0.26) in the label condition, 1.76 (SD = 0.25) in the ideogram condition, and 1.58 (SD = 0.47) in the intensity condition. An oneway ANOVA revealed no effect of condition, $F(2, 57) = 2.328$, $\eta^2 = 0.076$, p = .107, suggesting that

4 Analysis of all learner participants' data (*N = 84)* revealed qualitatively the same results, namely significant performance differences in the order: label > ideogram > intensity > no training (all three pairwise comparisons survived Bonferroni correction). Analysis of both learner and non-learner data (*N = 95)* revealed a similar—but not identical—gradation in performance across conditions: label > ideogram = intensity > no-training (significant differences surviving Bonferroni correction for three comparisons). This discrepancy may be attributed to the possibility that some of the seven non-learner participants in the intensity training condition were unable to disregard tone identity (as suggested by their informal reports). Thus, including non-learner data fails to test for the effect of cue individuation when exposure to stimuli is controlled.

participants' explicit knowledge of the highly predictive cues did not differ among training conditions.

To assess whether WPT accuracy was affected by explicit knowledge of the newly-trained names for the cues as inspected through the post-categorization questionnaire, we analyzed categorization data from the label training group only. A modified version of the mixed-effects model included a categorical fixed effect (with two levels, “No” and “Yes”) reflecting whether participants used the trained verbal labels in responding to the post-categorization questionnaire. This factor was not significant ($\beta = -.012$, $z = -.086$, $p = .932$) and did not interact with the other predictors (all $|\beta| < .003$, $p > .130$).

Correlation between training and categorization performance

Inspection of individual data revealed participants with high performance during training but low performance in the WPT, and vice versa. To investigate the possibility that cue training was predictive of subsequent categorization we regressed WPT performance onto average performance in the second half of Day 3 of training, potentially interacting with training condition. There was no significant effect of either condition or training performance and no significant interaction (all $p > .4$). Fig. 5 shows the scatter plot and the regression lines for the three training conditions as well as the regression line for data pooled from all three conditions. To explore the possibility that training performance was predictive of WPT performance depending on the number of cues forming a pattern, we separately calculated average WPT performance on 1-cue, 2-cue, and 3-cue trials. We regressed each performance measure onto average training performance in the second half of Day3, possibly interacting with training condition, and again there were no effects neither interactions for any of these analyses (all $p > .2$).

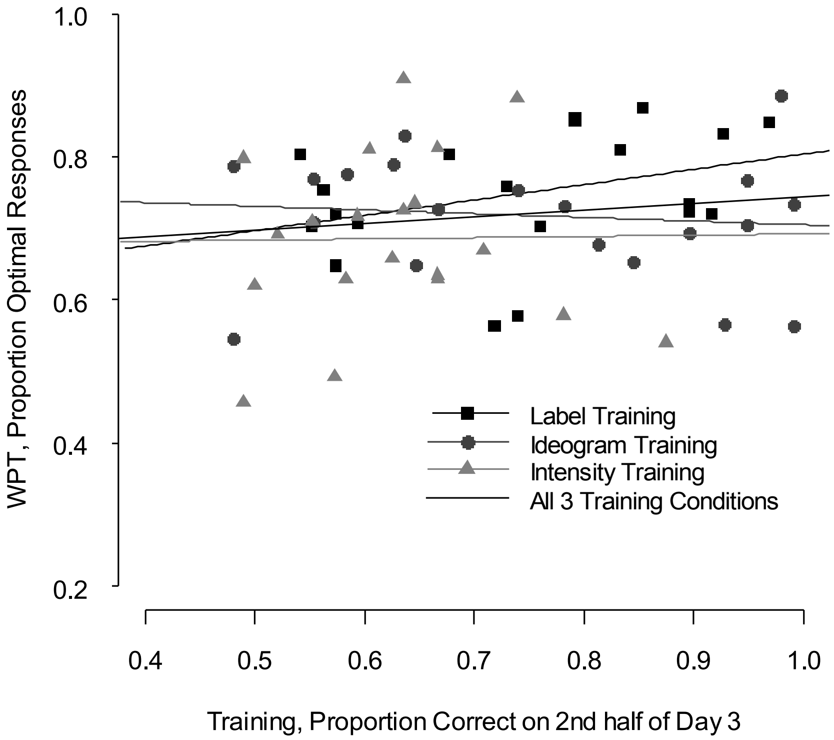


Figure 5. Scatter plot of WPT categorization performance versus training performance on the second half of Day 3. Lines correspond to linear regression parameter estimates.

Discussion

In this study, participants performed the WPT, a probabilistic category learning task, using hard-to-name auditory cues. In a training phase preceding the WPT, groups of participants learned to associate the cues to verbal labels or hard-to-name ideograms, or were exposed to the cues in an intensity task orthogonal to cue identity; there was also a group of participants receiving no training. Categorization performance in the WPT was significantly affected: the label training group outperformed the ideogram group, the ideogram training group outperformed the intensity training group, and the intensity training group outperformed the no-training group. Since all groups were administered the same auditory version of the WPT, the differences in performance can only be attributed to training. Therefore, (a) availability of verbal labels, (b) cue individuation, and (c) exposure to stimuli conferred independent benefits in the category learning task.

Verbal labels

We assumed that the availability of cue names would favor the formation, testing, and application of verbalizable strategies by participants in the label training condition because these participants would have easily accessible names for the cues of the categorization task. To ensure that availability of names was not confounded with categorical training, verbal label training was contrasted with ideogram training differing in the nonverbal nature of the associations. The advantage of the label training group suggests that cue names specifically enhanced explicit processes mediating WPT performance. The lack of significant differences in learning slopes between conditions further suggests that the naming advantage was not limited to early stages in WPT learning, perhaps serving simply as initial anchors, but extended throughout training. Also, participants' identification of the highly predictive cues, although a poor measure of awareness (see Lagnado et al., 2006, for a trial-by-trial assessment of task knowledge and self insight) suggests that awareness for the learned material was comparable among the training conditions, and thus precludes a potential explanation of the present results on the grounds of differential mediation of distinct memory systems in each condition.

Participants in the ideogram training group might have developed labels for the cues due to the extended exposure (cf. Galizio & Baron, 1976; Lupyan et al., 2007). Care was taken so that ideograms would be hard-to-name and that potential labels for the cues would not originate in them. Indeed, the post-training questionnaire confirmed the unavailability of easily accessible names for the ideograms, and the post-categorization questionnaire showed that very few participants in the ideogram training condition (4 out of 20) gave descriptions of the tones corresponding to ideograms' features. In contrast, in the label training condition, 11 out of 20 participants used the trained pseudowords to describe the tones, a significantly larger proportion ($\chi^2 = 3.84$, $df = 1$, $p = .05$). Even if labels were developed under ideogram training, the finding that the label group outperformed the ideogram group in the WPT—given equal performance at the end of training—suggests that these purported labels were largely idiosyncratic and ineffective.

It is conceivable that the advantage in categorization of the label group compared to the ideogram group might be due to more efficient encoding of the tones under label training. The difference in encoding efficiency might have resulted in a memory benefit (easier retrieval) when categorizing the tones. Identification of auditory warning sounds has shown more robust learning using verbal labels compared to “graphic” labels (Edworthy & Hards, 1999; though in some of the sounds graphic labels worked better and there were further confounds in that study). However, there is little reason to assume that auditory-verbal pairings resulted in an encoding advantage in our study, given our finding of equal training performance between the label and ideogram training groups at the end of training.

Another possible interpretation of the categorization advantage under label training would be increased perceptual discrimination of the cues (hypothesis of the “acquired distinctiveness of cues,” Miller & Dollard, 1941, as cited by Galizio & Baron, 1976). However, equal performance at the end of training in the label and ideogram conditions again argues against such an interpretation. Galizio and Baron (1976) suggested that acquired distinctiveness might be manifested with label training only when task conditions make cues difficult to discriminate. We have no reason to assume that the sequential presentation of the tones—with an interstimulus interval of 1 s—during the WPT imposes perceptual difficulty. Therefore the acquisition of perceptual features under label training does not seem to offer a strong explanation for our results.

It could be argued that label and ideogram group training differed in ways other than verbal labels. For example, the Chinese characters may be characterized by greater visual complexity than the printed pseudowords. This difference might not affect training but only manifest itself in a demanding task such as WPT. The present design cannot preclude this possibility, which must be explored in further research.

To explore the mechanisms that contributed to the difference in performance between the label and ideogram training groups we considered the possibility that WPT performance was driven by partial cue knowledge[5]. Given differences in training performance across participants and tones (e.g., not all participants were equally successful in learning the cue-response pairings for each of the four tones) we calculated each participant's individual cue knowledge, that is, the average performance for each of the four tones at the second half of the third day of training. Subsequently we constructed a measure of "partial cue knowledge" for each pattern and each participant in the WPT by averaging the participant's cue knowledge for the tones appearing in the pattern. This was only possible for participants in the label and ideogram training groups (because participants in intensity training did not classify tones by their identity). Data from the two conditions were re-analyzed with a modified mixed-effects model including partial cue knowledge (centered) as a fixed effect, along with its interactions. There was a four-way interaction involving target, trial, condition, and partial cue knowledge ($\beta = -.034$, $z = -3.124$, $p = .002$), hence data from the two conditions were separately analyzed. For the label training group there was a positive effect of partial cue knowledge on optimal responding (interaction of partial cue knowledge by target: $\beta = 1.722$, $z = 3.313$, $p < .001$), not interacting with trial (interaction of partial cue knowledge by trial and target: $\beta = -0.004$, $z = -0.511$, $p = .609$), consistent with a constant influence throughout the WPT. For the ideogram training group there was an interaction of partial cue knowledge by trial and target ($\beta = .030$, $z = 4.480$, $p < .001$) suggesting a variable effect of partial cue knowledge. Models with alternative trial centering revealed that partial cue knowledge had a negative effect in the first half of the procedure (e.g., at Trial 50, $\beta = -2.365$, $z = -3.124$, $p = .002$; at Trial = 100, $\beta = -.851$, $z = -2.191$, $p = .029$), no effect later on (at Trial = 150, $\beta = .633$, $p = .229$), and a positive effect at the end ($\beta = 2.144$, $z = 2.709$, $p = .007$).

This post-hoc analysis suggests that participants' categorization accuracy in the label training group was driven throughout the procedure by partial knowledge of the tone-label pairings. Participants performed better on those WPT trials that employed cues for which labels were better learned during training. This is consistent with the hypothesis that explicit hypothesis testing processes, mediated by the availability of verbal labels, are recruited

5 We thank an anonymous reviewer for suggesting this analysis.

during the WPT. Having names for the cues may have facilitated verbal working memory processes that contribute to category learning (Miles & Minda, 2011). In contrast, knowledge of tone-ideogram pairing seems to have interfered with WPT performance in the first half of the procedure. Perhaps the visual complexity of the ideograms distracted participants in the demanding WPT, impeding the formation of verbal, explicit rules. Further empirical investigation is needed to study this issue with planned comparisons in an appropriate design.

Cue individuation

The advantage in WPT performance of the ideogram training group compared to the intensity group may be attributed to the individuated representations formed for the tones during ideogram training. These representations, possibly akin to “perceptual anchors” (Ahissar, 2007), may have rendered the tones less abstract in working memory, thus facilitating the use of strategies when solving the WPT. In contrast, participants in intensity training could perform successfully disregarding tone identity, so task demands may not have caused the formation of individuated, concrete representations of the tones.

However, the ideogram and intensity training groups also differed in training performance, prior to WPT, leaving the WPT difference open to alternative interpretations that cannot be confidently rejected. For example, participants in the intensity training group may have recruited fewer or less efficient cognitive resources during training. The lower rate of successful performance produced diminished reinforcement—through positive feedback—and may have led to less efficient processing of the auditory tones. Further research with an easier training task is required to empirically assess this possibility.

The finding that cue individuation alone, in the absence of verbal labels, was beneficial to category learning in the WPT is important to the extent that the latter is primarily mediated by explicit processes, as it highlights the potential of individuated representations to participate flexibly in novel learning tasks. Previous research has suggested that cue characteristics are immaterial to WPT performance as long as there is an isomorphic probabilistic structure (Hopkins et al., 2004, Knowlton et al., 1994). In contrast, cue individuation seems to affect categorization performance, necessitating an explanation from memory systems approaches.

Prior exposure

Participants trained to associate sound intensity to colors exhibited greater categorization performance in the WPT compared to participants receiving no training at all. Notably, the intensity group was able to benefit from training explicitly requiring that the relevant dimension for later categorization (cue identity) be disregarded. The critical manipulation in this condition required participants to form intensity “categories” orthogonal to cue identity. Our pilot experiments showed that cue identity interfered with intensity judgments, so there is reason to hypothesize that cue identity and cue intensity are “integral” dimensions (Goldstone, 1994). On that account, it is possible that sensitization occurred along both dimensions during training and, thus, that intensity training enhanced discriminability among the cues (Goldstone, 1994). That this manipulation led to increased WPT performance compared to no training therefore suggests that (a) discriminability of the cues may be crucial for their effectiveness in probabilistic category learning and (b) exposure to stimuli is in itself beneficial for subsequent processing of these stimuli.

The beneficial effect of intensity training was especially apparent early in the WPT since participants in the no training condition exhibited near-chance performance in the first two blocks of 10 trials (see Fig. 4), reflecting perhaps an initial difficulty to identify the four tones. Generally, familiarity with the stimulus set is known to affect subsequent performance (e.g., Goldstone & Steyvers, 2001). More specifically, Folstein et al. (2010) exposed participants to artificial stimuli prior to a categorization task utilizing categorizing stimuli that were novel but had similar configuration as exposure stimuli. Even when the dimensions of exposure stimuli were uncorrelated and thus provided no diagnostic value for later categorization, there was a clear advantage in categorization performance compared to a group that remained unexposed to the stimuli. Perhaps participants were able to learn the structure of the stimuli and thus had an advantage in hypothesis testing or resource allocation. In our experiment participants received feedback for associating sound intensity to colors. However, the relevant dimension for training was absent in later categorization, similar to Folstein et al., allowing an explanation of the beneficial effect of exposure to stimuli in later categorization performance along the same lines.

Concerns and limitations

It is notable that average performance on the second half of Day 3 of training was not correlated to average WPT categorization performance for any of the training conditions. This may be interpreted as supporting the existence of discrete learning systems: training required gradual acquisition of cue-response pairings, whereas the WPT presumably required explicit hypothesis testing. At the moment, differences in task demands between the training and categorization task in our study do not allow us to draw firm conclusions in this matter (cf. Dunn & Kirsner, 2003). On the other hand, a more refined, by-cue measure of training performance was found to be predictive of between-trials differences in WPT performance . Partial knowledge of the cue-label pairings acquired during training was found to facilitate post-training categorization, whereas partial knowledge of the cue-ideogram pairings initially interfered with and later facilitated categorization. This connection between training and categorization provides no evidence in favor of a multiple systems account.

Observed differences in training performance between groups may cause some concern regarding the interpretations. Verbal label and ideogram training performance did not differ at the end of training, yet participants in the label training group probably achieved plateau performance (as evidenced by a lack of an effect of trial on Day 3) earlier compared to the ideogram training group (which kept on learning the cue-response pairings during Day 3, as evidenced by an effect of trial). We believe that this discrepancy between the two conditions does not pose a significant limitation on the interpretation of our results insofar as both groups' knowledge of the cue-response pairings was comparable at the end of the training procedure.

Another concern stems from the fact that the ideogram group outperformed the intensity group in training performance. Although similar performance in all three training conditions was desirable, the intensity training condition was primarily designed to equate exposure to the stimulus set and recruitment of attentional resources. The design constraint that tone identity be disregarded led to a significant difference in training performance at the end of training, leaving our results regarding individuation open to alternative interpretations.

Finally, we acknowledge that care should taken when interpreting the difference in WPT performance between the intensity and no-training groups. Participants in these conditions were—due to recruiting difficulties—sampled from different pools, hence no strong inferences can be made. This confound does not undermine the comparison of prime interest in our study, that is, between label and ideogram training.

Implications and conclusion

This is the first detailed report of gradual learning in an auditory version of the WPT. There are two procedural discrepancies between this version and the prototypical task (Knowlton et al., 1994), imposed by the auditory nature of the cues: First, the cues were presented sequentially, and second, feedback was delivered in the absence of the cues. There is evidence that both of these factors modulate the involvement of distinct memory systems during visual category learning (Foerde & Shohamy, 2011; Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005; Worthy, Markman, & Maddox, 2013; but see Dunn, Newell, & Kalish, 2012 for an alternative interpretation). However, it has been demonstrated that, for learning to take place, the appropriate mode of presentation for auditory stimuli is sequential and not concurrent, like in the visual modality (Conway & Christiansen, 2009; Saffran, 2002). Further research is required to examine if sequential presentation resulted in different memory system involvement compared to the prototypical WPT. Importantly, all our participants were administered the exact same version of the WPT. Therefore, procedural discrepancies between this auditory version and the prototypical WPT does not undermine the between-groups comparison that supports the idea that verbal labels facilitate explicit hypothesis testing.

The WPT has been extensively used as a tool by multiple systems (e.g., Knowlton et al., 1996; Poldrack and Foerde, 2008) and single system theorists (e.g., Newell et al., 2011) to assess the existence and relative contribution of discrete memory systems during categorization learning. It has been suggested that the majority of young healthy participants (Gluck et al., 2002; Poldrack & Foerde, 2008) initially approach the task by sub-optimal strategies that can be said to be declarative (Shohamy et al., 2008) but later on they engage multiple-cue (or integrative) strategies. These later strategies may be mediated by the procedural system (Shohamy et al., 2008) or they may be supported by declarative learning processes since they are accompanied by high levels of awareness (Price, 2009) or self-insight (Lagnado et al., 2006; Newell et al., 2007). Our results are consistent with the latter assumption. If the WPT is mediated by a procedural system and not by explicit hypothesis testing later in training, then having names for the cues should not affect later categorization performance. The fact that the label training group outperformed the ideogram training group throughout the task suggests that the declarative-procedural distinction does not explain healthy participants' behavior in the WPT. Instead, a general learning mechanism may support performance throughout the task (Newell et al., 2007).

To conclude, we have showed that newly trained verbal labels for the cues provide an advantage in probabilistic category learning performance. We based our hypothesis on the assumption that explicit hypothesis testing of verbal rules would be facilitated when having names for the cues, as opposed to associating the cues to difficult-to-name ideograms. The present results extend recent studies suggesting that language is not just for talking (Lupyan, 2008; Lupyan et al., 2007) and that verbal processes are important for categorization (Ashby & Maddox, 2005; Miles & Minda, 2011). Future research should examine in more detail the intuitive (but perhaps simplistic, see Newell et al., 2011) notion that humans may benefit from linguistic faculties during categorization with a new focus on verbal labels for categorizing items.

References

Abu-Shaba, N., Myers, C. E., Shohamy, D., & Gluck, M. A. (2001). *Age effects in probabilistic category learning*. Unpublished manuscript. Newark: Rutgers University.

Ashby, F. G, Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–81. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*, 147–161. doi:10.1111/j.1749-6632.2010.05874.x

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, England: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Bates, D., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes* (R package ver. 0.99875– 6). Retrieved from http://cran.rproject. org/web/packages/lme4/index.html

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.2.15, retrieved from http://www.praat.org/

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology, 21*, 561–580. doi:10.1080/09541440802097951

Dickerson, K. C., Li, J., & Delgado, M. R. (2011). Parallel contributions of distinct human memory systems during probabilistic learning. *NeuroImage, 55,* 266–276. doi:10.1016/j.neuroimage.2010.10.080

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59,* 447–456. doi:10.1016/j.jml.2007.11.004

Dunn, J. C. & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex, 39*, 1–7. doi: 10.1016/S0010-9452(08)70070-4

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 38, 840–859. doi:10.1037/a0027867

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Science, 103,* 11778–11783. doi:10.1073/pnas.0602659103

Foerde, K., Poldrack, R. A., Knowlton, B. J. (2007). Secondary task effects on classification learning. *Memory & Cognition, 35,* 864–874. doi:10.3758/BF03193461

Foerde, K. & Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience, 31*, 13157–13167. doi:10.1523/JNEUROSCI.2701-11.2011

Folstein, J., Palmeri, T. J., & Gauthier, I. (2010). Mere exposure alters category learning of novel objects. *Frontiers in Psychology, 1*, 1–6. doi:10.3389/fpsyg.2010.00040

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116–124. doi:10.3758/BF03195503

Fotiadis, F. A., Protopapas, A., & Vatakis, A. (2011). The effect of cue naming in probabilistic category learning. In: Kokinov, B., Karmiloff-Smith, A., & Nersessian, N. J. (Eds.), *European Perspectives on Cognitive Science: Proceedings of the European Conference on Cognitive Science – EuroCogSci 2011.* Sofia: New Bulgarian University Press.

Galizio, M., & Baron, A. (1976). Label training and auditory categorization. *Learning and Motivation, 7,* 591–602. doi:10.1016/0023-9690(76)90009-6

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247. doi:10.1037/0096-3445.117.3.227

Gluck, M. A, & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus, 3,* 491–516. doi:10.1002/hipo.450030410

Gluck, M. A., Shohamy, D., & Myers, C. E. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic classification learning. *Learning & Memory, 9,* 408–418. doi:10.1101/lm.45202

Goldstone R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General,* 123, 178–200. doi: 10.1037/0096-3445.123.2.178

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*, 116–139. doi:10.1037110096-3445.130.1.116

Holt, L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implication for first and second language acquisition. *Journal of the Acoustical Society of America, 119,* 3059–3071. doi:10.1121/1.2188377

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. A. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42,* 524–535. doi:10.1016/j.neuropsychologia.2003.09.005

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*, 150–166. doi:10.1016/j.jecp.2006.09.002

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

Knowlton, B. J., Mangels, J. A., Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399–1402. doi:10.1126/science.273.5280.1399

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1,* 106–120. doi:10.1101/lm.1.2.106

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General, 135,* 162–183. doi:10.1037/0096-3445.135.2.162

Lupyan, G. (2006). Labels facilitate learning of novel categories. In A. Cangelosi, A.D.M. Smith, & K. Smith (Eds.), *The Sixth International Conference on the Evolution of Language* (pp. 190–197). Singapore: World Scientific.

Lupyan, G. (2008). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137*, 348–369. doi: 10.1037/0096-3445.137.2.348

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18,* 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 650–662. doi:10.1037/0278-7393.29.4.650

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 100–107. doi:10.1037/0278- 7393.31.1.100

Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning & Memory, 13,* 230–239. doi:10.1101/lm.43006

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience and Biobehavioural Reviews, 32,* 237–248. doi:10.1016/j.neubiorev.2007.11.001

Miles, S. J., & Minda, J. P. (2011). The effect of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37*, 588–607. doi:10.1037/a0022309

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The Psychology of learning and motivation: Advances in research and theory, 52,* 117–162. San Diego: Academic Press. doi:10.1016/S0079-7421(10)52003-6

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences, 12*, 285–290. doi:10.1016/j.tics.2008.04.009.

Newell, B. R., Dunn, J. C., & Kalish, M. (2010), The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cogntion, 38,* 563–581. doi: 10.3758/MC.38.5.563

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed), *The Psychology of Learning & Motivation, 54,* 167–215. San Diego: Academic Press. doi:10.1016/B978-0-12-385527-5.00006-1

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Phychonomic Bulletin & Review, 14,* 505–511. doi:10.3758/BF03194098

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2,* 1–8. doi:10.3389/neuro.11.010.2008

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso, J., Myers, C. E., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550. doi:10.1038/35107080

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory system debate. *Neuroscience and Behavioural Reviews, 32,* 197–205. doi:10.1016/j.neubiorev.2007.07.007

Poldrack, R. A, & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiology of Learning and Memory, 82,* 324–332. doi:10.1016/j.nlm.2004.05.003

Price, A. L. (2009). Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task. *Memory & Cognition, 37,* 210–222. doi:10.3758/MC.37.2.210

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources & Evaluation, 46,* 449–459. doi:10.1007/s10579-010-9130-z

R Development Core Team. (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110,* 861–871. doi:10.1037/0735-7044.110.5.861

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 585–594. doi:10.1037/0278-7393.20.3.585

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196. doi: 10.1006/jmla.2001.2839

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioural Reviews, 32,* 219–236. doi:10.1016/j.neubiorev.2007.07.008

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118,* 676–686. doi:10.1037/0735-7044.118.4.676

Sturges, J., & Whitfield, T. W. A. (1995). Locating basic colours in the Munsell Space. *Color research and application, 20*, 364–376. doi:10.1002/col.5080200605

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82,* 171–177. doi:10.1016/j.nlm.2004.06.005

Tan, L. H., Chan, A. H. D., Kay, P., Khong, P.-L., Yip, L. K. C., & Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences, 105,* 4004–4009. doi:10.1073/pnas.0800055105

Van Casteren, M., & Davis, M. H. (2006) Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589. doi: 10.3758/BF03193889

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2013). Feedback and stimulus-offset timing effects in perceptual category learning. *Brain & Cognition, 81*, 283–293. doi:10.1016/j.bandc.2012.11.006

Yan, J., Qiu, Y., Zhu, Y., & Tong, S. (2010). Mental rotation differences between Chinese characters and English letters. *Neuroscience Letters, 479*, 146–151. doi:10.1016/j.neulet.2010.05.051

Yarkoni, T, Balota, D. A., & Yap, M. J. (2008). Moving Beyond Coltheart's N: A New Measure of Orthographic Similarity. *Psychonomic Bulletin & Review, 15,* 971–979. doi:10.3758/PBR.15.5.971

Supplementary Methods

Perceptual Equation of Tones

Each of the four tones were preprocessed in Praat (Boersma & Weenink, 2012), set to a reference intensity level, namely 10 dB below maximum. Ten participants (4 male, M_{age} = 31.6, *SD* = 2.72) who did not take part in the experimental conditions were administered 30 adjustment trials using PsychoPy (Peirce, 2009). In each trial two sounds were presented with an interstimulus interval of 500 ms. The first sound was always tone s1 in its reference intensity. The second sound was one of the three remaining tones presented in a pseudorandom intensity, half the times greater than the tone's reference intensity. Participants modified the intensity of the second sound in steps of 0.5 dB using two keys on the keyboard until they felt that the two sounds were equal in intensity. The three tones were adjusted 10 times each.

Points of subjective equality were calculated at +0.43 dB, +1.53 dB, and +3.00 dB for tones s2, s3, and s4, respectively. These intensity levels were called *adjusted* levels and were used in the training procedure for the 3 training groups.

Volume Calibration

Each participant heard the four tones at the 4 intensity levels (highest, high, low, and lowest, see Method), presented self-paced in random order, and adjusted the volume of their computer to a comfortable level. Subsequently, a 2AFC intensity discrimination task was administered, using 12 pairs of tones at different intensity levels with an interstimulus interval of 300 ms, presented at a pseudorandom order. Participants denoted the higher-intensity sound within 5 s. This procedure (volume adjustment and intensity discrimination) was repeated up to three times if necessary, to a criterion of no more than 3 discrimination errors (with a few exceptions). The computer's resulting volume setting was noted and participants were instructed to set the volume carefully before each training session. The procedure was programmed in DMDX (Forster & Forster, 2003).

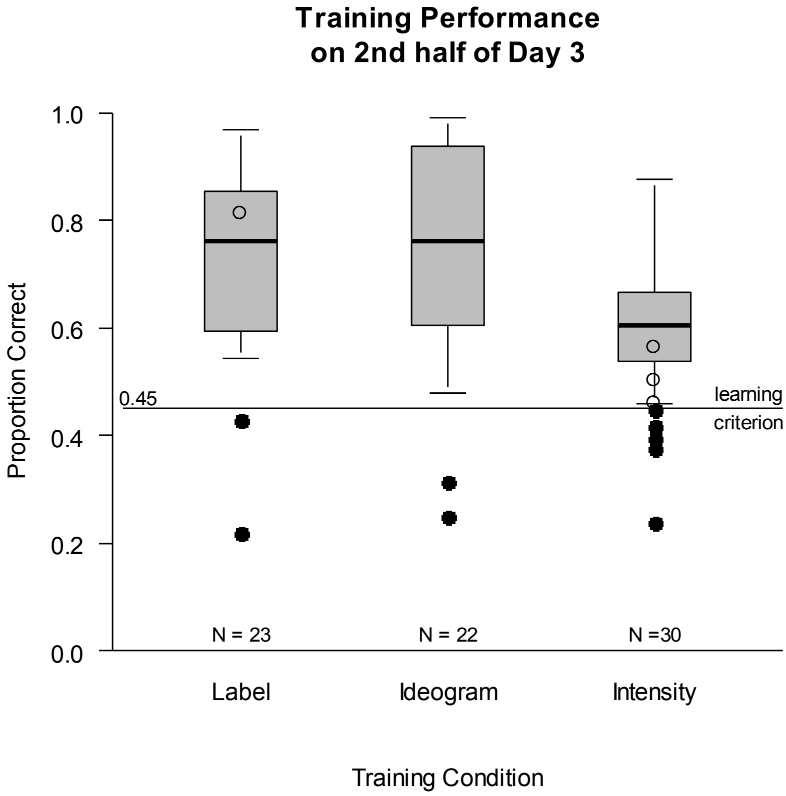


Figure S1. Average training performance on the last 96 trials of Day 3. Sample sizes on the graph denote the full number of participants in each condition. Bars enclose the middle 50% of individual performance, the median is marked with a thick line, and error bars extend to the full range of performance. Non-learner performance (below 45% correct, an arbitrary learning criterion) is marked with filled circles (*N* = 2 in label, *N* = 2 in ideogram, and *N* = 7 in intensity training condition). Open circles (*N* = 1 in label training, *N* = 3 in intensity training) depict performance of participants that were randomly excluded to form equal-sized samples (*N* = 20) in each condition.

References

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.2.15, retrieved from http://www.praat.org/

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116–124. doi:10.3758/BF03195503

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2,* 1–8. doi:10.3389/neuro.11.010.2008

3

No effect of verbal labels for the shapes on Type II categorization tasks

This chapter has been published as:

Fotiadis, F. A., & Protopapas, A. (2016). No effect of verbal labels for the shapes on Type II categorization tasks. In Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J.C. (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 277–282). Austin, TX: Cognitive Science Society

Abstract

Category learning is thought to be mediated—in at least some category structures—by hypothesis-testing processes. Verbal labels for the stimuli and stimulus individuation have been shown to facilitate the formation, testing, and application of rules of category membership (Fotiadis & Protopapas, 2014). We sought to replicate the phenomenon of facilitation due to verbal names for the stimuli by training participants for two consecutive days to either learn new names for abstract shapes, or learn shape-ideogram pairings; a third group was unexposed to the shapes. After training, participants were given a Type II categorization task—thought to be mediated by verbal processes of rule discovery—utilizing the trained shapes. We hypothesized that verbal labels for the shapes and shape individuation would provide facilitative effects in learning to categorize. Results revealed no effect of training on categorization performance. This study suggests that caution should be taken when generalizing findings across perceptual modalities or different experimental paradigms.

Introduction

The ability to categorize spans a broad range of human capacities and behaviors. Researchers have examined the cognitive processes (Ashby & Maddox, 2005) and neural substrates (Poldrack et al., 2001) of category learning and have utilized computational modeling techniques in an effort to shed light on the nature of the underlying representations (Anderson, 1991).

The Multiple Memory Systems (MMS) hypothesis argues that human category learning is mediated by distinct learning systems (Ashby & Maddox, 2005; Poldrack & Foerde, 2008). A declarative, explicit, or verbal system is thought to be engaged in the learning of categories that can be characterized by a verbal rule. Hypothesis testing processes are thought to be recruited, and the knowledge acquired is thought to be available to consciousness. On the other hand, the learning of categories that defy a simple verbal description is thought to be accomplished through a procedural, implicit, or non-verbal system. Pre-decisional perceptual processes underlie learning, and the learned material is thought to be unavailable to consciousness. An on-going debate exists between the MMS theorists and single-system theorists arguing that a single, general learning mechanism suffices to account for behavioral data (e.g., Newell, Dunn, & Kalish, 2011).

In the context of this debate, growing empirical evidence suggests that verbal processes are important in the learning of rule-described categories. Ashby and colleagues (Ashby, Alonso-Reese, Turken, & Waldron, 1998) developed a computational theory suggesting that the verbal system mediates rule-based category learning. Verbal working memory interference has been found to impair the learning of rule-described categories (Miles & Minda, 2011), whereas experimental manipulations, such as using difficult-to-name stimuli (Kurtz, Levering, Stanton, Romero, & Morris, 2013), or verbal rehearsal of stimulus dimensions prior to learning (Minda, Desroches, & Church, 2008), have been shown to affect category learning.

Verbal Labels in Hypothesis Testing

Recently, Fotiadis and Protopapas (2014) provided evidence in favor of the hypothesis that verbal labels for the to-be-categorized stimuli facilitate hypothesis-testing processes underlying category learning. The authors utilized hard-to-name auditory stimuli and manipulated the availability of stimulus names by training separate groups of participants for three consecutive days to associate the auditory tones with pseudowords (label training condition) or with hard-to-name ideograms (ideogram training condition); or to associate tone intensity with colors (intensity training condition); a fourth group remained unexposed to the tones (no-training condition). On the fourth day all participants were administered the same auditory version of the Weather Prediction Task (Knowlton, Squire, & Gluck, 1994) utilizing the trained tones as cues. Results revealed a gradation in categorization performance in the order: label > ideogram > intensity > no-training. Thus, it was concluded that verbal labels, cue individuation, and exposure to the stimulus set each facilitated explicit hypothesis-testing processes underlying category learning.

The Shepard et al. (1961) Tasks

In their seminal paper, Shepard, Hovland and Jenkins (1961) revolutionized the study of category learning. They created six category tasks (Type I to Type VI) by manipulating category structure (categorization rule) while utilizing the same stimuli in each task and the same number of exemplars in each category. In the most common implementation of the paradigm (Minda & Miles, 2010) categorization stimuli are comprised of three binary valued dimensions: Shape (square vs. triangle), color (black vs. white), and size (big vs. small).
The basic finding of the Shepard et al. (1961) study was that the order of difficulty of the six types (as assessed by participants' performance) cannot be accounted for by a simple stimulus-generalization theory. The key finding was that participants found it easier to learn Type II categories compared to Type IV categories, despite the reduced within-category similarity of the former (compared to the latter) category structure. The authors suggested that this Type II over Type IV advantage necessitates considering the mediation of executive attention mechanisms and the formulation and application of rules during category learning.

The Type II Task and Rule-Discovery Learning

The Type II task has a two-dimensional rule structure. Two out of the three dimensions are diagnostic of category membership , in an exclusive-or fashion. A simple verbal rule seems to be able to define category membership[6] (e.g., “black triangles and

6 Depending on which dimensions are diagnostic, there can be three Type II subtypes: Shape-irrelevant, size-irrelevant, and color-irrelevant. See Love and Markman (2003) for evidence suggesting that performance varies systematically across these subtypes.

white squares are category A”). Thus, the structures' processing demands are thought to be best met by explicit rule-learning processes (Minda & Miles, 2010).

This claim seems to be supported by empirical evidence. Minda et al. (2008) utilized the first four prototypical Shepard et al. (1961) category structures in an effort to examine rule-selection executive functions of children and adults. In their Experiment 2, Minda et al. assessed the effect of a concurrent verbal and a concurrent non-verbal task on categorization performance. The verbal secondary task—thought to occupy resources recruited by verbal processes of rule discovery—did impair performance in the Type II structure (compared to a control, no-task condition, and also compared to the non-verbal task condition). These results suggest that the Type II structure recruited the explicit system. Smith, Minda and Washburn (2004) studied category learning processes of human and non-human animals using the Shepard et al. tasks. Their results provided evidence in favor of the engagement of rule-discovery mechanisms in learning the Type II category structure. The evidence (“all-or-none learning”) was only present for human subjects, whereas for non-human animals, lacking the faculty of language, there was no sign of rule discovery. This may be considered as evidence in favor of the engagement of rule-learning mechanisms in the Type II task.

Thus, theoretical reasons (Minda & Miles, 2010; Shepard et al., 1961) as well as empirical evidence (Minda et al., 2008; Smith et al., 2004) suggest that learning to categorize in the Type II category task is mediated by hypothesis testing processes of verbal rules.

Rationale of the Present Study

The purpose of the present study was to further test the hypothesis that verbal labels for the to-be-categorized stimuli facilitate hypothesis testing processes recruited during category learning (Fotiadis & Protopapas, 2014). We specifically wanted to test our training manipulation in the visual modality, since there are reasons to suggest that learning across modalities is not governed by the same mechanisms (Conway & Christiansen, 2005). Moreover, given that in the Weather Prediction Task category membership is probabilistically defined (Knowlton et al., 1994) we sought to examine the effect of names for the stimuli in a task with a deterministic structure.

To manipulate the availability of names, separate groups of participants were trained for two consecutive days to associate abstract shapes with pseudowords (label training condition) or with hard-to-name ideograms (ideogram training). A third—control—group of participants remained unexposed to the shapes, and was trained to associate ideograms with pseudowords (mock training condition). On the second day, all participants were given the Type II categorization task. In this task, the two values in the shape dimension were the same shapes that were used in the training procedure. The category-diagnostic dimensions were shape and color, whereas size was non-diagnostic (see Fig. 1).

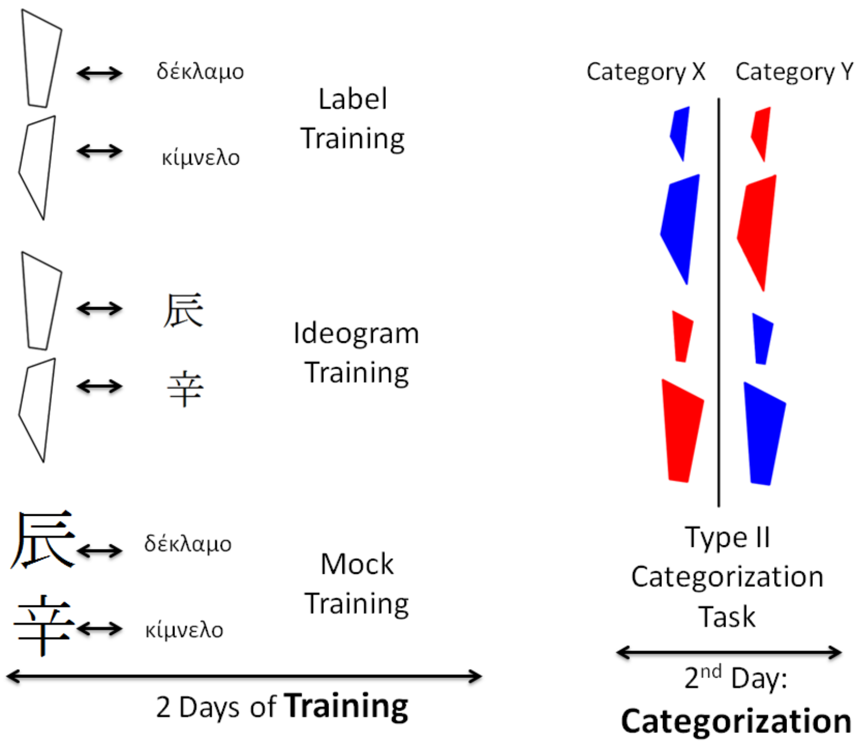


Figure 1. Design of the present study.

We reasoned that verbal labels for the shapes would facilitate verbal hypothesis-testing processes of rule formation, testing, and application. Therefore we predicted that participants in the label training condition would find it easier to discover the categorization rule, compared to the ideogram training group. We also hypothesized that familiarity with the stimuli and learning to associate the shapes to visual stimuli (ideograms) would help create individuated perceptual representations of the shapes and therefore facilitate categorization. We therefore predicted that the ideogram training group would have an advantage in rule discovery compared to the mock training group.

Methods

Participants

Seventy-two students (16 male) of the University of Athens took part in the study and were randomly assigned (in groups of 24) to each training condition. Their mean age was 25.8 years (*SD* = 7.0). All were native speakers of Greek, reported normal or corrected-to-normal vision, no history of neurological illness, and no diagnosis of dyslexia.

Materials

Shapes. Two abstract shapes of low association value were selected from the collection of Vanderplas and Garvin (1959). The shapes have been previously used in experimental research and are considered to be hard to name (e.g., Hulme, Goetz, Gooch, Adams, & Snowling, 2007). The shapes were equated in size in a pilot experiment using the method of adjustment. Twelve participants took part in this psychophysical procedure, which was implemented in PsychoPy (Peirce, 2007), and the results provided the Points of Subjective Equality (P.S.E.s). For the training session, empty shapes with a black margin were created, with size corresponding to 75% of the P.S.E.s, whereas for the categorization session the shapes were filled with red or blue color. The categorization stimuli necessitated two levels of size for the stimuli, so for the “big” shapes the size corresponded to the P.S.E.s, and the “small” shapes were created by a 50% reduction in size.

Pseudowords. Ten pseudowords were created, equated in number of letters, syllables, phonemes, and stress position. They were also roughly equated in orthographic and phonological typicality using Levenstein distance of the 20 nearest neighbors (Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012; Yarkoni, Balota, & Yap, 2008). To avoid name assignment biasing toward particular shapes, we administered an online questionnaire to 107 native speakers of Greek showing randomly one of the two abstract shapes along with the ten candidate pseudowords. Participants were simply asked to “choose a name” for the shape. We selected the two pseudowords that were selected as names for both shapes with roughly equal frequency, namely δέκλαμο (/'ðɛklamo) and κίμνελο (/'kimnɛlo).

Ideograms. Two Chinese characters were selected. These ideograms have been previously used and have been shown to resist a simple verbal description (Fotiadis & Protopapas, 2014): 辛 (U+8F9B), and 辰 (U+8FB0). To equate for number of strokes (and, thus, for perceptual complexity), a stroke was erased from the second character.

Procedure

Training comprised two sessions, administered on two consecutive days. On the second day, following training, the categorization task was administered. All following procedures were implemented in the DMDX display software (Forster & Forster, 2003).

Training. There were two training sessions, administered on consecutive days, aimed to allow overnight consolidation. Participants were given 160 trials in each training session, arranged in four blocks of 40 trials. At the beginning of a Label Training trial a fixation cross was presented for 500 ms at the center of the screen. Following that, one of the two shapes was randomly selected and presented for 2000 ms, and then the two pseudowords appeared in a vertical configuration. Participants were asked to respond by clicking on one of the two alternative responses (pseudowords). Upon response, feedback was provided (the word "Correct" or "Wrong") for 500 ms. The permutation of the two pseudowords was counterbalanced across trials, and each shape was presented equally often within a block of trials. On the first day of training, participants in the Label training condition were asked to read aloud the pseudoword of their choice before clicking on it, because we reasoned that learning a name necessitates the formation of an effective phonological component. This reading-aloud instruction was omitted on the second day, to equate task demands across training conditions as much as possible.

For the Ideogram Training condition pseudowords were replaced with ideograms. In the Mock Training condition participants were asked to learn to associate ideograms to pseudowords, so shapes were replaced by ideograms. No reading aloud took place in either the Ideogram or the Mock training conditions. Stimulus-response pairings (e.g., shape-pseudoword or shape-ideogram pairings) were counterbalanced across participants within a training condition. Each training session lasted approximately 20 minutes.

Categorization. In the Categorization session, which followed immediately after the second training session, participants were told that they would learn to classify stimuli into two categories, namely X and Y. They received a maximum number of 28 blocks of eight trials. In a categorization block each of the eight categorization stimuli was presented once. The beginning of a trial was signaled by the presentation of the stimulus at the center of the screen along with the category labels X and Y. The category labels were presented around the stimulus either horizontally or vertically in both permutations (i.e., X - Y or Y - X), providing four possible configurations. Participants responded by clicking on a category label. Following each response a smiling face was presented if correct or a frowning face if incorrect, for 1500 ms. If a participant did not respond within 10000 ms the trial was terminated and a prompt appeared on screen. The training session ended upon completion of all blocks, or upon two consecutive errorless blocks (following Mathy, Haladjian, Laurent, & Goldstone, 2013). The order of trials was pseudorandomized with MIX (VanCasteren & Davis, 2006) and was identical for all participants. Randomization constraints precluded (a) presentation of the same categorization stimulus on two consecutive trials, and (b) a lag between trials with the same configuration of category labels less than two. The assignment of category label (X-Y) to values of the diagnostic dimensions was counterbalanced across participants. A short break was provided every 56 trials. The maximum duration of the categorization session was 25 minutes.

Results

Training

Participants in all three training conditions mastered the training task by the third block of Day 1 and exhibited ceiling performance on Day 2. Across both training sessions, participants averaged 98.33% correct responses (*SD* = 1.16) in the label training condition, 97.33% (*SD* = 2.64) in the ideogram training condition, and 97.87% (S*D*= 2.29) in the mock training condition. The average performance per training condition and day, in blocks of 40 trials, is shown in Fig. 2.

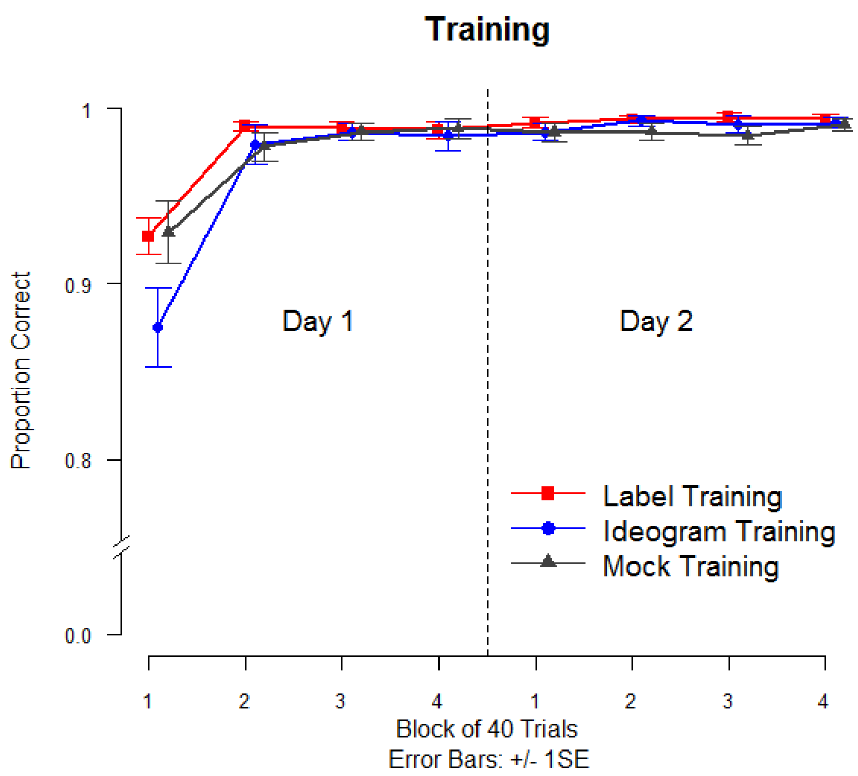


Figure 2. Learning curves in the three training conditions in blocks of 40 trials

The purpose of the analysis was to test if participants were equally successful in learning the shape-label and shape-ideogram pairings. Therefore, we only analyzed data from the label and ideogram training conditions on the second day of training. Participants' responses were analyzed in R (R Development Core Team, 2014) with a linear mixed-effects model including fixed effects of trial and training condition, as well as their interaction, and random effects of participants. By-participant random slopes of trial were included to model participants' individual learning rates (Baayen, Davidson, & Bates, 2008). In R notation, the model was specified as

```
accuracy ~ trial*condition+(1+trial|participant).
```

There was no simple effect of trial (β = .387, z = 1.317, p = .188), a result consistent with participants' ceiling performance on Day 2. There was also no interaction of trial by condition (β = −.031, z = −.113, p = .910) and—most importantly—no effect of condition (β = .369, z = .692, p = .489).

Categorization

Out of 72 participants in all training conditions, 44 (61.11%) managed to achieve two consecutive errorless blocks of trials, thus providing unequivocal evidence of having discovered the categorization rule. We refer to these participants as "learners." There were 14 learners (58.33%) in the label training condition, 14 learners (58.33%) in the ideogram training conditions, and 16 learners (66.67%) in the mock training condition. A chi square test revealed that the percentage of learners in the categorization task did not differ significantly between training conditions ($\chi2$ = .468, *df* = 2, *N* = 72, *p* = .792).

However, using the percentage of participants reaching the learning criterion as a dependent variable has the disadvantage of disregarding the ease or difficulty with which participants in each training condition learned the rule. We therefore analyzed the number of blocks to reach criterion, for learner participants only[7] (see Fig. 3). An one-way analysis of variance revealed that there was no effect of training condition on the number of blocks to reach the learning criterion, $F(2, 42) = 1.777$, $\eta^2 = .080$, $p = .182$.

7 The exclusion of "non-learner" data is common practice in the categorization literature when analyzing number of blocks to reach criterion (e.g., Mathy & Feldman, 2009). The rationale is that a value of 28 corresponding to a non-learner, perhaps responding at chance, and a value of 28 corresponding to a participant mastering the task at the last two blocks reflect qualitatively different behaviors that should not be aggregated.

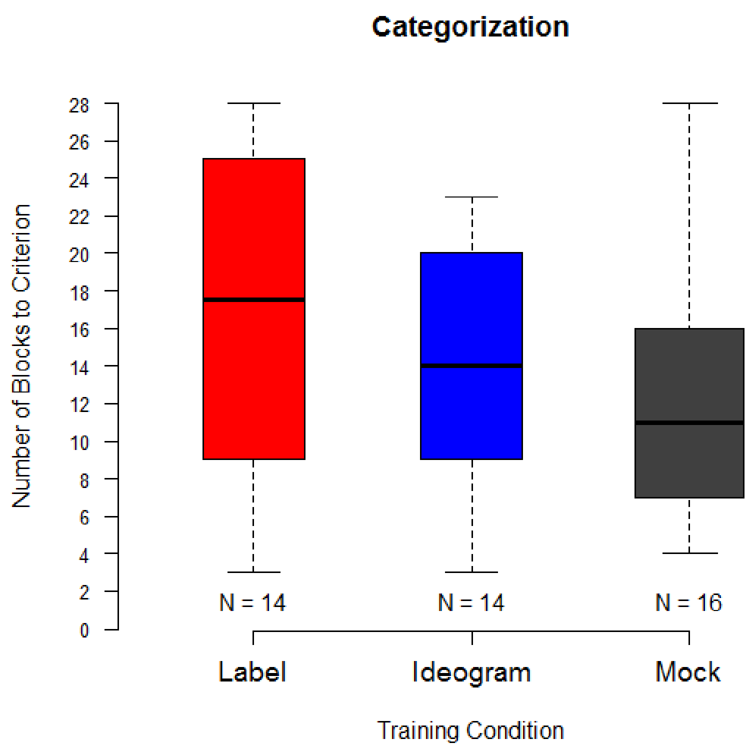


Figure 3. Number of blocks to reach learning criterion in the categorization task, per training condition. Data from learner participants only. Boxes denote interquartile range; thick lines mark the median; error bars extend to the full range; N denotes sample size.

Alternatively, categorization performance can be analyzed using accuracy as the dependent variable. This allows inclusion of all participants, under the assumption of errorless performance after the learning criterion is reached (e.g., Kurtz et al., 2013). A linear mixed-effects model with the same formula as above revealed an effect of trial (β = 4.878, z = 5.33, p < .001), reflecting an increase in accuracy as trials progressed, comparable learning rates among conditions (all βs < 1.35, p > .14), and—most importantly—no effect of condition on categorization accuracy (all βs < .13, p > .32).

Discussion

In this study we trained participants for two consecutive days to learn new names for shapes, or learn to associate shapes with hard-to-name ideograms. A third group of participants remained unexposed to the shapes. In a categorization task, administered immediately after training, we used the trained shapes to create the categorization stimuli. We predicted that names and familiarity with the shapes would each facilitate rule discovery in the categorization task. Our results revealed no effect of training condition of categorization, in contrast to previous findings (Fotiadis & Protopapas, 2014).

This discrepancy raises concerns about assuming that an effect manifesting itself in one modality would also be present in another modality. One purpose of the experiment was to replicate the effect of facilitation in learning to categorize due to names for the stimuli in the visual modality. The lack of an effect may be attributed to the change in modality per se, since there is reason to assume that learning processes may differ between modalities. Saffran (2002) showed that, for learning to take place, the temporal mode of presentation of stimuli in the visual and auditory modality should be different (concurrent vs. sequential respectively). Also, Conway and Christiansen (2005) implemented the same learning paradigm in different modalities and provided evidence in favor of a learning advantage in the auditory modality compared to the visual modality. Further empirical investigation is needed to assess whether learning to categorize in the auditory and the visual modality is mediated by the same processes.

Participants' ceiling performance during training complicates interpretation, insofar as potential differences between learning the shape-label and shape-ideogram pairings may be masked by the ease of the task. Thus, we cannot preclude the possibility that performance in the categorization task is affected by differences in training.

Alternatively, the lack of an effect of verbal labels in category learning may stem from methodological discrepancies between the present and our previous study, such as the structure of the categorization task used to reveal hypothesis learning processes. The Weather Prediction Task, previously shown to be affected by names for the stimuli, has a probabilistic structure, whereas the Type II task used in the present study has a deterministic structure. It remains to be investigated whether performance in a probabilistic category structure may be more easily affected by experimental manipulations, perhaps due to the uncertainty that is inherent in the task.

Further concerns stemming from the results of the present study are related to whether changing the surface structure of a paradigm affects the processing demands of a task. The result of no difference in categorization performance between the label and ideogram training groups might suggest that names for the stimuli do not facilitate rule discovery. An alternative explanation, however, may be related to the fact that our implementation of the Type II task utilized two abstract shapes whereas the canonical version uses two geometric shapes. It may be that the Type II task is learned through verbal processes of rule discovery only when the values of the diagnostic dimensions are highly familiar to participants. Mathy et al. (2013), who also used abstract shapes in implementing the Type II task, provided evidence in favor of the engagement of similarity-based processes (thought to reflect learning mediated by the implicit rather than the rule-based system) in learning to categorize. Thus, although the Type II task has been used to examine explicit processes (e.g., Minda & Miles, 2010), our version of the task may have recruited implicit processes that are not affected by verbal labels for the stimuli.

A final concern may be of representational nature. The finding that familiarity with the stimuli also failed to affect performance in the categorization task seems rather puzzling, given previous findings and current understanding in the field. For example, Folstein, Gaultier, and Palmeri (2010) provided evidence suggesting that mere exposure to the stimulus configuration may facilitate subsequent categorization performance. Our finding of no significant difference in performance between the ideogram and mock training conditions may be taken to indicate that learning processes involved in learning to categorize our version of the Type II task did not recruit the representations of the shapes that were presumably acquired during training. Indeed, informal reports of participants' strategies in debriefing revealed that participants mainly paid attention to the corners of the shapes and not to the shape forms in their entirety. Therefore, a plausible explanation for our findings is that the participants learned names for the entire shapes and formed individuated representations of them but then only used parts of the shapes in the categorization task. The representational mismatch undermined the potential of the verbal labels and the familiarity with the shapes to facilitate learning in the categorization task.

To conclude, we sought to replicate the effect of facilitation in learning a verbal rule of category membership caused by having names for the stimuli. The results suggest that learning processes may operate differently across modalities or across categorization paradigms and that task processing demands may be significantly altered if the surface structure of a categorization paradigm is modified.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24–39.

Folstein, J., Gauthier, I, & Palmeri, T. J. (2010). Mere exposure alters category learning of novel objects. *Frontiers in Psychology, 1*, 1–6.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124

Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory & Cognition, 42*(1), 112–125.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*(2), 150–166.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106–120.

Kurtz, K. J., Levering, K., Romero, J., Stanton, R. D., & Morris, S. N. (2013). Human learning of elemental category structures: Revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 552–572.

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition, 31*(5), 790–799.

Mathy, F., Haladjian, H. H., Laurent, E., & Goldstone, R. L. (2013). Similarity-dissimilarity competition in disjunctive classification tasks. *Frontiers in Psychology, 4*(26), 1–14.

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review, 16*(6), 1050–1057.

Miles, S. J., & Minda, J. P. (2011). The effect of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 588–607.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and nonrule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1518–1533.

Minda, J. P., & Miles, S. J. (2010). *The influence of verbal and nonverbal processing on category learning*. In B. H. Ross (Ed.), The Psychology of learning and motivation: Advances in research and theory, 52, 117–162. San Diego: Academic Press.

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). *Systems of category learning: Fact or fantasy?* In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 54, pp. 167–215). San Diego: Academic Press.

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso, J., Myers, C. E., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory system debate. *Neuroscience and Biobehavioral Reviews, 32*, 197–205.

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources and Evaluation, 46*(3), 449–459.

R Development Core Team. (2014). R*: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in Rhesus Monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General, 133*(3), 398–414.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*(4), 584–589.

Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology, 57*, 147–154.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971–979.

4

Category labels facilitate learning and affect subsequent processing of category-diagnostic perceptual features

This chapter has been submitted for publication

Introduction

Category learning is instrumental for survival and ubiquitous in everyday life. Human and nonhuman animals (Smith, Minda, & Washborn, 2004) form categories from a very early age (Waxman and Markow, 1995) and researchers have used a variety of techniques, such as mathematical modeling (Nosofsky, 1986), fMRI (Poldrack et al., 2001) and eye tracking (Rehder & Hoffman, 2005a; 2005b), to investigate the mechanisms and cognitive processes underlying the learning of categories.

Language has been suggested to play a formative role in human cognition and not just serve communicative purposes (for a review see, e.g., Gleitman & Papafragou, 2013). In particular, in the field of category learning, verbal labels for the learned categories have been argued to influence categorization processes. Evidence to support this link originates in developmental psychology, where it has been shown that children's formation of categories is affected by correlated linguistic cues (Landau, Smith, & Jones, 1988; Yoshida & Smith, 2005). Likewise, children benefit in categorization when categories are accompanied by verbal labels (Waxman & Markow, 1995) and this benefit is associated with verbal labels but not other cues (such as tones; Fulkerson & Waxman, 2007).

The facilitative effect of verbal labels for the categories is also evident in adults with fully developed linguistic capacities. Specifically, Lupyan, Rakison and McClelland (2007) trained participants to classify figures of alien creatures in two categories. Following each categorization decision a redundant verbal label was presented. Results suggested that verbal labels (either visual or auditory) for the categories facilitated learning, compared to non-verbal (location) cues or to the absence of cues. Thus, accumulating evidence suggests that categorization processes are affected by language, and in particular by verbal labels for the categories.

The Label-Feedback Hypothesis

The label-feedback hypothesis was postulated by Lupyan (2012a; 2012b) to explain phenomena of categorical grouping on the processing of categorization items. According to this theory, perceptual, categorical, and linguistic representations are not encapsulated but, rather, interact. Building on Goldstone's (1994) work on the warping of perceptual space due to category learning, Lupyan suggested that labels for the categories exert an influence on perceptual representations of categorization items by selectively activating perceptual features that are diagnostic for categorization. This influence results in more “prototypical” (Lupyan, 2012b) or “categorical” (Lupyan, 2012a) perceptual representations, in the sense that perceptual differences that are important for categorization are emphasized whereas unimportant differences are deemphasized.

The activation of diagnostic features due to labels is not an all-or-none phenomenon but, rather, depends on the level of activation of the verbal labels. Lupyan (2012a; 2012b) suggested that the effect of labels may be manipulated through the up- or down-regulation of participants’ linguistic activity. For example, an up-regulating manipulation may be the overt presentation of labels for well-practiced categories (“concepts”) at the beginning of an experiment (Lupyan & Spivey, 2008) or at the beginning of each experimental trial (Lupyan & Thompson-Schill, 2012). Conversely, linguistic activity may be down-regulated through transcranial direct current stimulation (tDCS) over Wernicke’s area (Perry & Lupyan, 2014) or through verbal interference (Lupyan, 2009; see Perry & Lupyan, 2013, for a critical review of such methodologies). Thus the effect of category labels on shaping perceptual representations was assumed to be both pervasive and dynamically controlled.

The label-feedback hypothesis has received support from studies assessing the effect of labels for well known (“overlearned;” Lupyan & Spivey, 2010b) categories using visual search tasks (Lupyan & Spivey, 2008), same–different discrimination (Lupyan, 2008b; Lupyan, Thompson-Schill, & Swingley, 2010), picture verification (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), probe detection (Lupyan & Spivey, 2010b),“odd-one-out” procedures (Lupyan, 2009), object detection (Lupyan & Spivey,2010a; Lupyan & Ward, 2013), and object recognition (Lupyan, 2008a). This research program has mainly recruited adult healthy participants (but also aphasic patients; Lupyan & Mirman, 2013). Experimental methodologies have also included eye tracking (Edmiston & Lupyan, 2015), tDCS (Perry & Lupyan, 2014), and electroencephalography (Boutonnet & Lupyan, 2015). Thus, the effect of verbal labels for overlearned concepts might be said to be well supported. But what about learning novel artificial categories?

The Label Advantage During The Learning Of Novel Categories

According to the label-feedback hypothesis, participants in the label conditions of the Lupyan et al. (2007) study outperformed those in the control conditions because of perceptual space warping over the course of learning, as labels selectively activate diagnostic perceptual features (Lupyan, 2012a). However, a review of the literature on label advantages during category learning suggests that the phenomenon is not ubiquitous. Brojde, Porter and Colunga (2011) found no facilitation in the label condition of Lupyan et al. (2007) after changing the diagnostic dimensions to shape or texture and shape/hue or brightness. In one case they found slowing down of category learning instead. Using a similar procedure, Tolins and Colunga (2015) found no label advantage during learning to categorize. Perry & Lupyan (2014) found no effect of up-regulation (through overt redundant labels) or down-regulation (through tDCS) of linguistic activity on categorization performance. Finally, Lupyan and Casasanto (2015) contrasted the effects of different redundant verbal labels on categorization performance. They found an advantage from redundant real words, compared to the no-label condition, as well as from pseudowords, but only when they were selected to activate the same class of perceptual features as the words.

In sum, the facilitative effect of labels during learning novel categories (Lupyan et al., 2007) has been shown to be absent (Brojde et al., 2011; Perry & Lupyan, 2014; Tolins & Colunga, 2015), reversed (Brojde et al.), or replicable only under special selection of experimental materials (Lupyan & Casasanto, 2015). This constitutes a challenge for the label-feedback hypothesis and warrants further examination of the replicability of the phenomenon, which we undertake in Experiment 1.

Long-Lasting Effects

An important question concerns whether learning novel categories under verbal labels has long-term effects on subsequent tasks (Tolins & Colunga, 2015). There are several reasons to expect such sustained effects. The first line of evidence comes from perceptual learning. Goldstone (1994) showed that the learning of categories results in the warping of perceptual space; most notably, the category-diagnostic perceptual dimension is sensitized. This has been documented in tasks immediately following categorization training (Folstein, Gauthier, & Palmeri, 2012; Folstein, Palmeri, & Gauthier, 2013; Folstein, Palmeri, & Gauthier, 2014; Goldstone, 1994; Goldstone & Steyvers, 2001; Van Gulick & Gauthier, 2014) or administered after several days (Folstein, Palmeri, Van Gulick, & Gauthier, 2015). Thus, the evidence supports the notion of “stable dimension modulation,” that is, a long-lasting effect of perceptual space warping due to category learning (Folstein et al., 2015).

Therefore, if labels for the categories result in selective activation or sensitization of their diagnostic features, this could have lasting effects: The diagnostic features of named categories should be activated to a greater extent than the diagnostic features of hard-to-name categories in tasks following initial training, because this activation has been "learned." Selective activation may result in increased saliency, which serves to capture attention. We therefore assumed that in postcategorization tasks the diagnostic features of categories linked to verbal labels should capture attention to a greater extent compared to the diagnostic features of categories linked to other types of cues.

Converging evidence in favor of long-lasting category label effects come from studies of selective attention, thought to underlie the learning of categories (e.g., Nosofsky, 1986; Shepard, Hovland, & Jenkins, 1961). In a categorization task involving a change in the diagnosticity of perceptual dimensions, Kruschke (1996) found that participants learned to pay attention to the diagnostic dimensions early in training, and perseverated in subsequent processing. Similarly, Goldstone and Steyvers (2001) manipulated dimension diagnosticity and found that selective attention is shaped during the learning of categories and transferred to subsequent categorization tasks with either beneficial or detrimental effects on performance.

These studies are relevant to the label-feedback hypothesis to the extent that labels for the categories have been shown to affect attention during learning. Brojde et al. (2011) and Perry and Lupyan (2014) suggested that, when the learning of categories may be based on more than one dimensions, redundant verbal labels shift attention to dimensions typically known to be predictive of category membership. Moreover, Perry and Lupyan (2016) suggested that selective attention may be thought of as the warping of perceptual space, influenced by verbal labels for the categories. Therefore, if labels affect attention during category learning (Brojde et al; Perry and Lupyan, 2014), then attention should be shifted to the features that are diagnostic of named categories to a greater extent compared to features that are diagnostic of hard-to-name categories. Based on research suggesting that attention is learned (Goldstone & Steyvers, 2001; Krushke, 1996), we assumed that this effect would be evident in posttraining tasks.

According to the label-feedback hypothesis (Lupyan, 2012a; 2012b) the learned labels are predicted to activate the diagnostic features of named categories to a greater extent compared to the diagnostic features of hard-to-name categories. This differential activation is again assumed to affect attention so that the diagnostic features of labeled categories will capture attention to a greater extent compared to the diagnostic features of hard-to-name categories.

All three accounts—namely, perceptual learning, learned attention, and the label-feedback hypothesis—lead to the expectation of long-term effects of category labels, despite their qualitative differences in the origin and nature of the predicted effect. Thus, inspired by the literature on category learning, we examined sustained effects of category labels on subsequent processing in Experiment 2, using post-training categorization (e.g., Goldstone & Styevers, 2001; Krushke, 1996), and in Experiment 3, using a visual discrimination task (e.g., Goldstone, 1994) while monitoring participants' eye movements.

The Initial and Sustained Effects of Labels on Paired-Associate Learning

A question that has yet to be addressed is whether the selective activation of diagnostic features due to verbal labels is specific to categorization. The label-feedback hypothesis postulated an activation mechanism (Lupyan, 2012a; 2012b). This was not a general mechanism underlying learning under any condition but, rather, was specifically postulated to account for phenomena of categorical perception (Lupyan, 2012a) and the effect of conceptual grouping on visual and cognitive processing (Lupyan, 2008a; 2008b). Therefore, the effects of this mechanism (both initial and sustained) should be observed only when learning to categorize. This hypothesis was examined in Experiments 2 and 3 by comparison to control learning conditions, in which the initial training task required participants to form named and hard-to-name associations instead of categories, while the procedure and learning material remained largely the same.

Verbal stimuli have been argued to be processed more efficiently compared to hard-to-name stimuli, since verbal labels may serve as material symbols (Tolins & Colunga, 2015) that stabilize ideas in working memory (Lupyan et al., 2007). Therefore, an initial effect of labels, namely a label advantage, may also appear when learning to associate, due to a general facilitation in processing verbal stimuli. On top of this general facilitation, the mechanism of selective activation of diagnostic features during learning to categorize is expected to provide additional facilitation, specific to category learning. Therefore the label advantage during learning to categorize was predicted to be greater compared to the label facilitation during learning to associate.

With respect to the sustained effects of labels, no evidence neither a hypothesis has been offered regarding sustained effects of learning verbal compared to hard-to-name associations. Therefore, following paired-associate learning, no difference is expected in the posttraining processing of diagnostic features of named compared to hard-to-name associations.

Contrasting category and paired-associate learning has been utilized in the past by Poldrack et al. (2001) and has proven fruitful in illuminating categorization processes. However, there were important differences between category and paired-associate training in the Poldrack et al. study (e.g., response presentation and feedback delivery), which limit the generalizability of conclusions that can be drawn and necessitate a novel methodological approach with improved control over the features of the two training regimes. Contrasting the effect of labels on learning to categorize with the corresponding effects on learning to form associations is arguably a strict test of the label-feedback hypothesis (Lupyan, 2012a) that has not been previously implemented. Before embarking on this endeavor, however, it is instrumental to examine if the label advantage during learning to categorize (Lupyan et al., 2007) is replicable; this we undertake first in Experiment 1.

辰

辛

辰

辛

辰

辛

辰

辛

辰

辛

Figure 1. Design of Experiments 1, 2, and 3. (A) Participants in Experiment 1 learned four new artificial categories. Two of the categories were denoted by verbal labels, whereas the other two were denoted by hard-to-name visual symbols. (B) In the training phase of Experiment 2, a group of participants learned named and hard-to-name categories (in a procedure identical to Experiment 1). A separate group of participants learned named and hard-to-name associations. Following training there was a categorization session: Both groups were administered three category-learning tasks. The first (control) task utilized novel shapes and was meant to serve as a rule-discovery task. The last two tasks had the same formal structure as the control task (shape and color were the category-diagnostic dimensions) but utilized the shapes that had previously been predictive of either named or hard-to-name categories (or associations, depending on training condition). (C) Experiment 3 had two phases: In the training session, separate groups of participants learned either named and hard-to-name categories or named and hard-to-name associations (in a procedure identical to the training session of Experiment 2). Following training all participants were administered an eyetracking visual discrimination task employing the previously trained shapes.

Experiment 1

Participants in Experiment 1 were trained to learn four novel artificial categories. Two of the categories were denoted by verbal labels, whereas the other two categories were denoted by visual symbols (Chinese ideograms; see Figure 1A) that were previously found to resist a common verbal description (Fotiadis & Protopapas, 2014). We hypothesized that hard-to-name symbols would down-regulate linguistic activity (see also Kurtz, Levering, Romero, Stanton, & Morris, 2013, for a similar argument), leading to less efficient learning in the ideogram categories compared to the verbal label categories (Lupyan et al., 2007). Such a facilitative effects of labels can only be observed if the two shapes that are predictive of the label categories are activated to a greater extent compared to the two shapes that are predictive of the ideogram categories. This is consistent with previous research showing that categorization sensitizes a perceptual dimension *selectively*, facilitating the processing of specific values/instances, rather than generally (Goldstone, 1994; Van Gulick & Gauthier, 2014; for more on this issue, see the General Discussion).

Method

Participants. Twenty-four students (two male) of the University of Athens took part in exchange for course credit, and their mean age was 19.7 (*SD* = 1.1) years. All participants reported normal or corrected-to-normal vision, Greek as their native language, and no diagnosis of dyslexia.

Materials.

Categorization items. Four four-point abstract shapes of low association value (and thus considered hard-to-name; Hulme, Goetz, Gooch, Adams, & Snowling, 2007; MacLeod & Dunbar, 1988) from the Vanderplas and Garvin (1959) repository were perceptually equated in size, using the method of adjustment (implemented in PsychoPy; Peirce, 2007) to obtain points of subjective equality (PSEs). 288 categorization items were created—72 for each shape—by varying the size (randomly within 0.2–0.8 of the PSE) and border color (randomly selected hues) of the PSEs.

***Pseudowords*.** Two pseudowords served as response cues for the label categories, namely σάβης (/'savis/) and ρήτης (/'ritis/), previously used by Fotiadis and Protopapas, (2014). The two pseudowords were equal in numbers of letters, syllables, and phonemes, stress position, and orthographic typicality (the mean orthographic Levenshtein distance of the 20 nearest neighbors—OLD20—was 2.00 for both pseudowords taking stress into account; Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012; Yarkoni, Balota, & Yap, 2008).

***Ideograms*.** Response cues for the ideogram categories were two hard-to-name Chinese characters (previously used by Fotiadis & Protopapas, 2014), namely 辛 (U+8F9B) and 辰 (U+8FB0). One stroke was erased from the second character to equate number of strokes—and thus perceptual complexity.

Procedure. Participants were administered 288 training trials, organized in 12 blocks of 24 trials. Each shape was presented equally often within a block. Participants never saw a categorization item twice.

Participants were told that they would be presented with four different shapes in varying size and color, and with four responses: two names and two ideograms. Their job was to learn which shape (disregarding color and size) corresponded to each response.

At each trial a fixation cross was presented for 500 ms, followed by a categorization item presented for 2000 ms. The two pseudowords and the two ideograms appeared next in a random vertical arrangement for a maximum of 10000 ms. Participants responded by clicking on a response option and feedback—the words “correct” or “wrong” in Greek—was delivered for 500 ms. The procedure was programmed in DMDX display software (Forster & Forster, 2003). Participants were given eight practice trials in the beginning and a short break after every four blocks. The task lasted approximately 35 minutes.

The order of categorization items and the permutation of response cues were pseudorandom (implemented with MIX; Van Casteren & Davis, 2006), with constrains precluding the same permutation of response cues in consecutive trials, and the same shape in more than two consecutive trials. All possible permutations of response cues were presented in each block. Participants received the same order of categorization items. Assignment of shapes to categories was counterbalanced across participants with the constraint that the shapes were paired, so that two shapes belonging to a pair were both predictive of either label or ideogram categories. This resulted in eight possible combinations of shape-response assignment, with three participants randomly assigned to each combination.

Data Analysis. Analyses were conducted in R version 3.3.3 (R Core Team, 2017), employing generalized additive mixed models (Wood, 2011) with binomial distributions (Dixon, 2008), via a logit tranformation (Jaeger, 2008), fitted with restricted maximum likelihood and marginal likelihood estimation using package mgcv (Wood, 2011). Model comparison and visualization of model estimates was done using package itsadug (Van Rij, Wieling, Baayen, & van Rijn, 2016).

Results

Trials in which participants did not respond and trials with response latencies shorter than 250 ms were excluded from analyses (eleven trials total, 0.16 % of the data). Participants gradually improved in learning both the label and ideogram categories, averaging overall 81.9% (SD = 11.9) correct responses. Their accuracy was 83.0% (SD = 13.6) in the label categories and 80.8% (SD = 11.2) in the ideogram categories. Learning curves, in blocks of 24 trials, can be seen in Figure 2A.

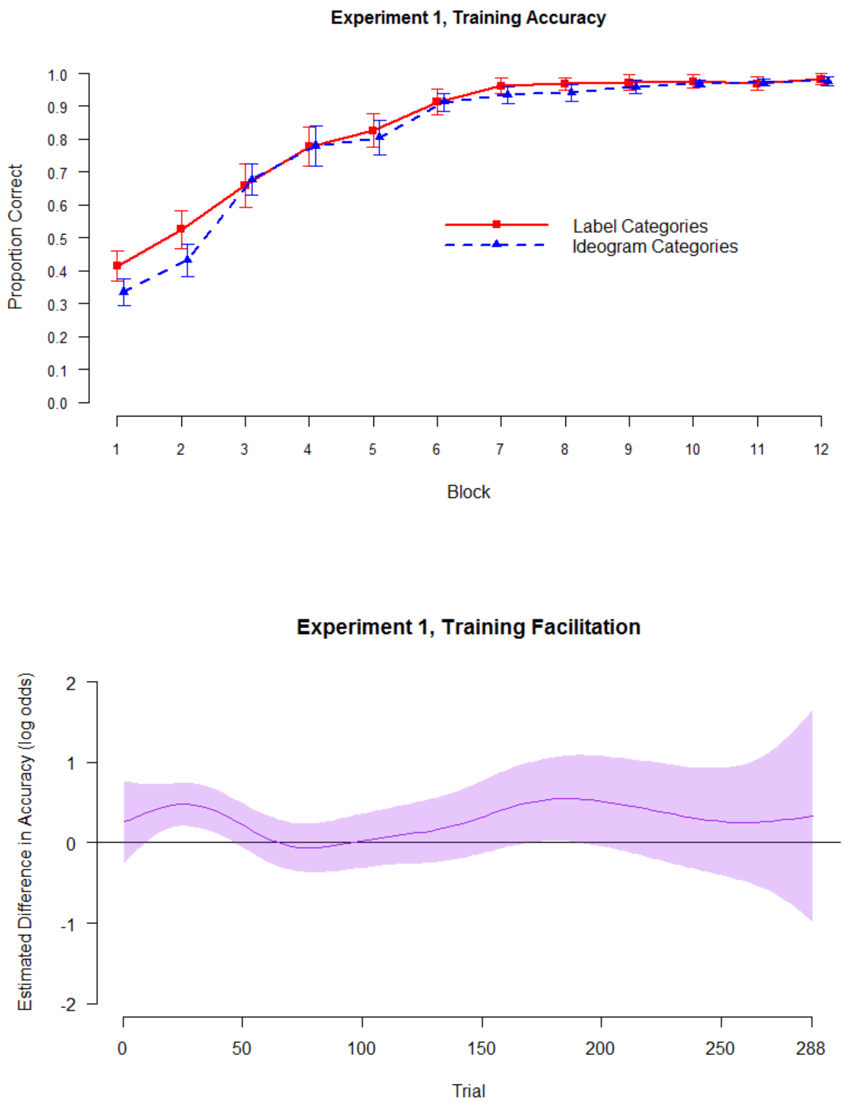


Figure 2. Results of Experiment 1. (A) Training accuracy of the label and ideogram categories in blocks of 24 trials. Error bars show between-subjects standard errors of the means. (B) Estimated difference in accuracy (in log odds) between the label and ideogram categories excluding random effects of participants. Error bands show 95% confidence intervals of the estimates.

Accuracy (a binomial variable) was regressed on condition (a within-subjects factor with levels “label” and “ideogram”) and trial (1–288, centered and scaled to *SD* = 1). Model comparison procedures led to the most parsimonious model with two smooth terms of trial (one for the label and one for the ideogram categories) and by-participants random smooth terms of trial to capture individual longitudinal variability. In R notation, the best-fitting model was:

```
acc ~ 1 + condition + s(trial, by = condition) + s(trial,
subject, bs = "fs", m = 1),
```

with s(trial, by = condition) denoting two smooth terms of trial, and s(trial, sbj, bs = "fs", m = 1) denoting by-participant random smooth terms of trial.

The question of whether participants' performance in the two conditions differed, and if so at which trials, was addressed through visualization (Baayen, 2013). Figure 2B shows the 95% confidence interval of the model's estimate for the difference in accuracy (in log odds) between the label and ideogram categories, excluding random effects. Participants exhibited increased accuracy in learning the label compared to the ideogram categories for two periods in the task, in particular for values of trial in the range [10, 45] and also in the range [172, 190] (the values were back-transformed from the normalized values in the model).

Discussion

Experiment 1 provided support to the label-feedback hypothesis (Lupyan, 2012a; 2012b) by revealing a label advantage during learning to categorize: Participants were more accurate in learning the label categories compared to the ideogram categories (Lupyan et al., 2007) for two periods in the task. To the best of our knowledge, this is the first time the label advantage is found in a within-subjects experiment, supporting the assumption that following category learning a perceptual dimension is selectively sensitized (Goldstone, 1994).

Experiment 2

Experiment 2 was designed to examine the long-term effects of labels for newly-learned categories on attention processes recruited during the processing of category-diagnostic features. As detailed in the Introduction, we hypothesized that shapes that had previously been predictive of named categories (hereafter “label shapes”) would capture attention to a greater extent compared to shapes that had previously been predictive of hard-to-name categories (“ideogram shapes”) during posttraining test tasks. To test this, three Type II (Shepard et al., 1961) category learning tasks were administered immediately following the training session (see Figure 1B).

Categorization items in the Type II category structure are composed of three binary-valued perceptual dimensions. In the most common implementation (Minda & Miles, 2010) items are simple geometric stimuli differing in shape (e.g., triangles vs. squares), size (e.g., big vs. small), and color (e.g., black vs. white). Two of the dimensions are diagnostic of category membership in an exclusive-or fashion: For example, black triangles and white squares belong to category A. Participants learn to ignore the size dimension and pay attention to shape and color, therefore attention processes are considered instrumental in learning to solve to the Type II tasks (Krushke, 1992; 1996; Nosofsky, Palmeri, McKinley, & Glauthier, 1994; Nosofsky & Palmeri, 1996; Rehder & Hoffman, 2005a).

In the second phase of our experiment, we manipulated the values of the shape dimension, using the label shapes in one Type II task (“label task”) and the ideogram shapes in the other Type II task (“ideogram task”) (see Figure 1B). We hypothesized that the label shapes would capture attention to a greater extent compared to the ideogram shapes, hindering deployment of attention to other dimensions. Therefore, paying attention to both shape and color, as was required for successfully solving the Type II tasks, was assumed to be more difficult in the label task compared to the ideogram task. Therefore, a sustained effect of labels for the categories is predicted to lead to differing accuracy in learning to categorize in the two Type II tasks, despite their identical formal structure.

To deter the involvement of rule-discovery processes in the two experimental—label and ideogram—tasks (Minda, Desroches, & Church, 2008; Minda & Miles, 2010), a control Type II (Shepard et al., 1961) task was administered first to each participant, before the two critical Type II tasks that involved the previously trained (label and ideogram) shapes. If participants have already learned, during this control task, that it is the shape and color of categorization items that define category membership in an exclusive-or fashion, they should be able to apply the same solution in the two subsequent tasks (see also Levine's, 1975, transfer hypothesis). A learning criterion was therefore introduced in the control task, further examining the performance of only “learner” participants, to ensure that the functioning of attention (rather than rule-discovery) processes is contrasted during learning to categorize in the two experimental tasks.

A second issue was also addressed in Experiment 2, contrasting the effects of labels for categories—both initial and sustained—with the effects of labels for associations. Accordingly, a group of participants was administered paired-associate training instead of category learning in the first phase of the experiment, followed by the same three Type II (Shepard et al., 1961) tasks in the second phase (see Figure 1B). The mechanism of selective activation of diagnostic features due to verbal labels was assumed to be categorization-specific. We therefore predicted that the initial effect of category labels, namely the label advantage during learning to categorize, would be of greater magnitude compared to a purported label advantage during learning to associate. With respect to the sustained effects of labels over the Type II tasks (in the second phase), we have no reason to assume that learning associations would cause attention to be differentially captured by shapes paired with either verbal or hard-to-name response cues. Thus, participants in the paired-associate group were predicted to exhibit comparable accuracy in categorizing the label and ideogram Type II tasks.

Method

The structure of the experimental design is shown in Figure 1B. Briefly, in the first phase two groups of participants learned to categorize, or to associate, pairs of shapes with either verbal labels or ideograms as response cues. In the second phase all participants had to solve three successive Type II tasks with the exact same structure. The last two of these involved the same shapes as the first-phase learning tasks, to test for the sustained effects of labels.

The category-training task was identical to Experiment 1. Thus, only the paired-associate-training task and the categorization session will be described here.

Participants. Overall 88 students took part in Experiment 2, meeting the requirements for normal or corrected-to-normal vision, no diagnosis of dyslexia, and Greek being their native language. A sampling-with-replacement method was followed, to achieve the design counterbalancing with participants meeting the control-Type II-task learning criterion, so that data are reported below from 64 learner students, including 32 (five male) assigned to the category-training group (age *M* = 21.8 years, *SD* = 4.6) and 32 (12 male) assigned to the paired-associate-training group (*M* = 23.2, *SD* = 6.3).

Materials.

***Paired-associate training*.** Four association items were created, identical to the abstract shapes used in category training but with black margin and size corresponding to 75% of the PSEs (see Figure 1B). The same pseudowords and ideograms used in category training were used as response cues, making this task a close parallel of Experiment 1 category training, differing only in the number of different stimuli associated with each response cue (many in category training vs. one in paired-associate training).

***Categorization session*.** An additional pair of four-point abstract shapes from Vanderplas and Garvin (1959) was used for the control Type II (Shepard et al., 1961) task (see Figure 1B). These novel shapes had also been equated in size along with the abstract shapes of Experiment 1. The size of the "big" categorization items corresponded to the PSEs, whereas "small" items were set to 60% of the PSE size. The color of categorization items in the control task was light blue and brown, whereas in the label and ideogram tasks color was either red and green, or blue and yellow. Overall, 40 categorization items were created (eight for the control task, 16 for the label task, and 16 for the ideogram task).

Procedure.

***Paired-associate training*.** The paired-associate-training task was a modification of the category-training task in that (a) every categorization item was replaced with the corresponding association item, and (b) instructions made no reference to either shape or size.

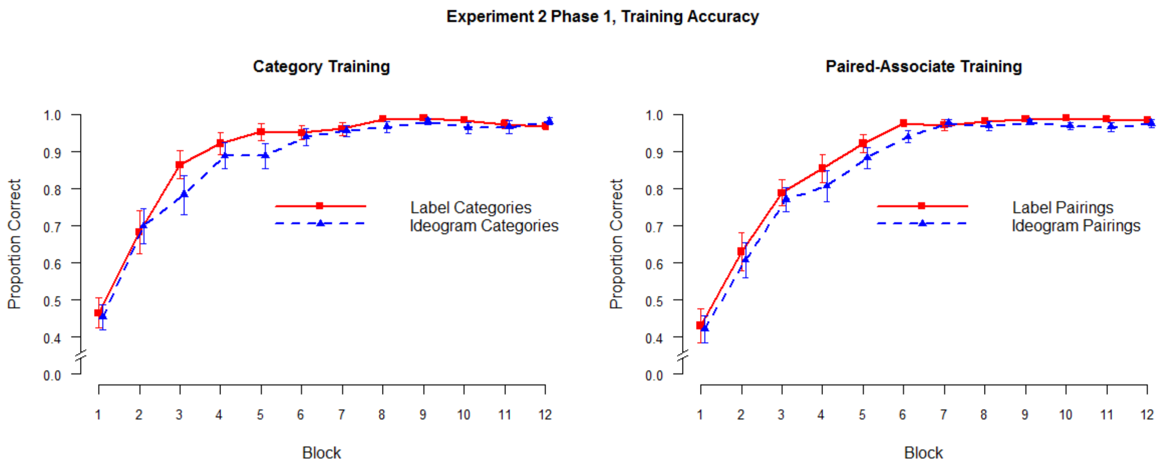
Categorization session. The three Type II (Shepard et al., 1961) tasks only differed with respect to categorization items and category labels. To avoid effects unrelated to our hypothesis, dimension diagnosticity (shown to affect learning; Love & Markman, 2003) was not counterbalanced but, rather, shape and color were the two diagnostic dimensions in all three tasks for all participants. Color and task order for the label and ideogram tasks was counterbalanced across participants.

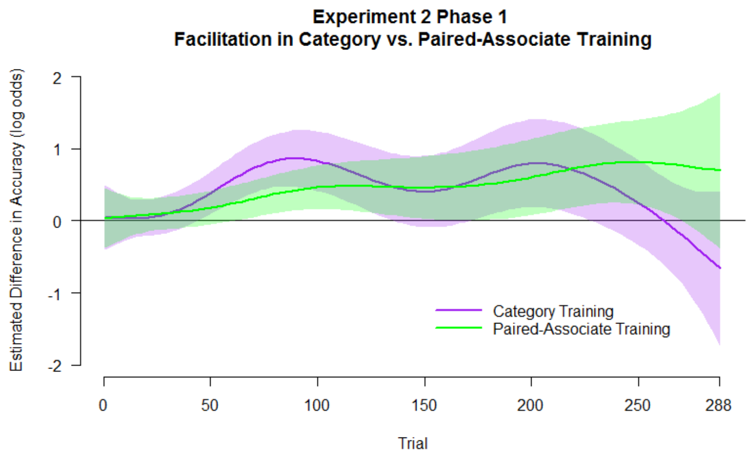
Participants received a maximum of 32 blocks of eight trials in each task. In each block all eight categorization items were presented in random order. Unbeknownst to participants, the task ended upon completion of all trials or upon achievement of the learning criterion, that is, two consecutive errorless blocks of trials (following, e.g., Kurtz et al., 2013; Mathy, Haladjian, Laurent, & Goldstone, 2013).

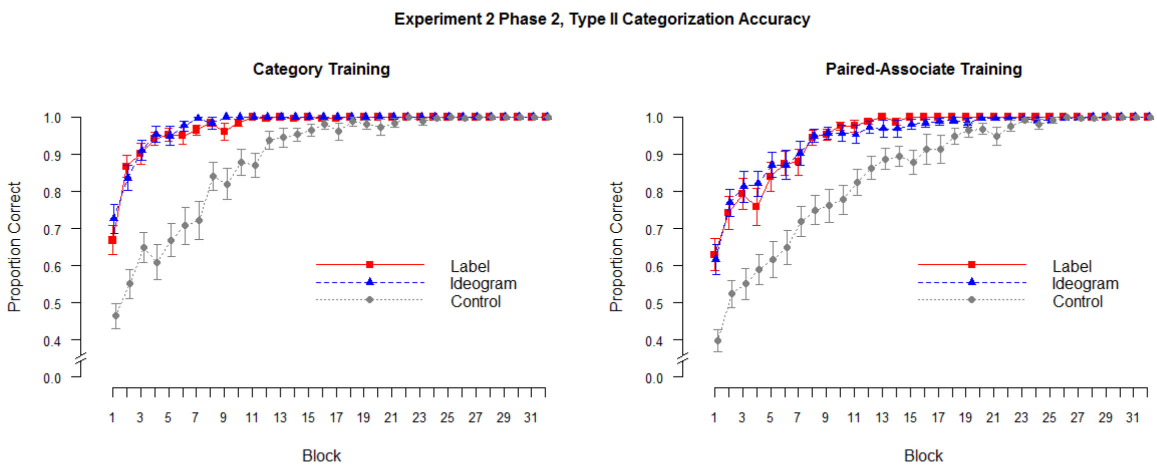
Before each task, participants were told that in each trial they would be presented with an item belonging to one of two categories. Their job was to find the categorization rule and make correct classification decisions. They were also informed that the stimuli would be in one of two shapes, one of two colors, and either big or small size, and they were instructed to be as accurate as possible.

A trial was initiated by the presentation of a categorization item along with the category labels (e.g., “1” and “2”) on its left and right side for a maximum of 20 s. Participants responded by pressing one of two keys on the keyboard. Upon response, the categorization item disappeared and feedback was presented for 1500 ms, consisting of a smiley face and high-pitched tone for correct response, or a frowning face and low-pitched tone for incorrect. In case of no response, a prompt appeared urging participants to make a key press. The intertrial interval was 0 ms. As in Experiment 1, assignment of shape pairs to label or ideogram response cues was counterbalanced across participants. This resulted in 32 combinations for each training group, to which participants were randomly assigned. Participants were given eight practice trials before each task and a short break every 7 blocks. The maximum duration of the categorization session was approximately 75 minutes.

Data analysis. To facilitate the analysis of categorization accuracy in the Type II tasks we followed the assumption of errorless performance after the learning criterion (Nosofsky et al., 1994; Nosofsky & Palmeri, 1996; see Kurtz et al., 2013, for experimental verification). Analyses were conducted with generalized additive mixed models using packages mgcv (Wood, 2011) and itsadug (Van Rij et al., 2016) in R version 3.3.3 (R Core Team, 2017).


Experiment 2 Phase 1, Training Accuracy
Category Training
Paired-Associate Training
Proportion Correct
Block
Label Categories
Ideogram Categories
Label Pairings
Ideogram Pairings


Experiment 2 Phase 1
Facilitation in Category vs. Paired-Associate Training
Estimated Difference in Accuracy (log odds)
Trial
Category Training
Paired-Associate Training


Experiment 2 Phase 2, Type II Categorization Accuracy
Category Training
Paired-Associate Training
Proportion Correct
Block
Label
Ideogram
Control

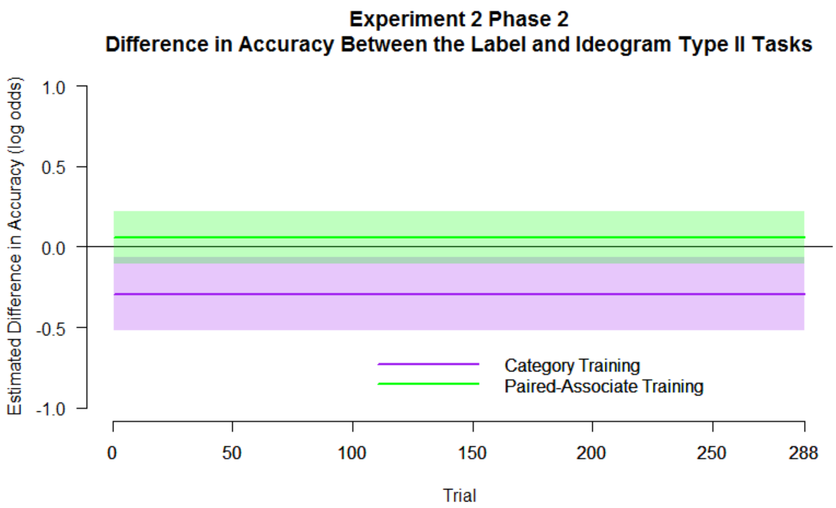


Figure 3. Results of Experiment 2. (A) Training accuracy of the label and ideogram categories (left panel) and of the label and ideogram pairings (right panel) in blocks of 24 trials. Error bars show between-subjects standard errors of the means. (B) Estimated difference in accuracy (in log odds) between the label and ideogram categories and the label and ideogram pairings, excluding random effects of participants. Error bands show 95% confidence intervals of the estimates. (C) Categorization accuracy in the label, ideogram, and control Type II tasks, for the category-training group (left panel) and the paired-associate-training group (right panel) in blocks of eight trials. Error bars show between-subject standard errors of the means. (D) Estimated difference in categorization accuracy (in log odds) between the label and ideogram tasks for the category- and the paired-associate-training groups excluding random effects of participants. Error bands show 95% confidence intervals of the estimates.

Results

Training session. No-response trials and trials with latency less than 250 ms were excluded from analyses (4 trials, comprising 0.04% of the data, from the category-training group, and 11 trials, comprising 0.12% of the data, from the paired-associate-training group). Participants' accuracy in learning both the label and ideogram categories increased as trials progressed, averaging 88.2% (*SD* = 9.0) correct over all 288 trials in the category-training group (*M* = 89.2%, *SD* = 9.3, in the label categories, and *M* = 87.2%, *SD* = 9.9, in the ideogram categories) and 86.6% (*SD* = 6.9) in the paired-associate-training group (*M* = 87.5%, *SD* = 7.9, in the label pairings, and *M* = 85.6%, *SD* = 7.4, in the ideogram pairings). Figure 3A depicts average learning curves for the category- and the paired-associate-training group, in blocks of 24 trials.

A dummy variable was defined to encode the interaction of training group by response cue ("igr": Interaction of Group by Response) in four distinct levels, namely category-labels, category-ideograms, paired-labels, and paired-ideograms. Trial was normalized, as in Experiment 1. Model comparison procedures suggested that the most parsimonious model was the one with different smooth terms of trial for each level of the group×response interaction variable. Individual variation in performance was modeled with by-participants random smooth terms of trial. In R notation, the best-fitting model was:

```
acc ~ igr + s(trial, by = igr) + s(trial, sbj, bs =
"fs", m=1),
```

with s(trial, by = igr) denoting four smooth terms of trial, and s(trial, sbj, bs = "fs", m = 1) denoting by-participant random smooth terms of trial.

The purpose of this analysis was to assess if labels facilitated category learning more than associative learning. Figure 3B plots the model's estimates of the label facilitation in the category- and the paired-associate-training group, in log odds, excluding random effects. In category learning, accuracy was greater in learning the label compared to the ideogram categories for values of trial (back-transformed) in the range [44.31, 134.34] and also in the range [172.09, 227.26]. In comparison, in paired-associate learning, accuracy was greater in learning the label compared to the ideogram pairings for values of trial in the range [61.73, 267.92]. Thus facilitation due to labels was revealed during the learning of both categories and associations. The 95% confidence intervals of the estimated facilitation in the two groups overlap throughout the task (see Figure 3B), so the label advantage was statistically indistinguishable across the learning of both categories and associations.

Categorization session. To facilitate analysis, accuracy was set to 1 in (nonadministered) trials following two consecutive errorless blocks. Three no-response trials were excluded from analyses (0.01% of the data, all from the paired-associate-training group). As shown in Table 1, participants in both groups exhibited increased accuracy in the two tasks following the initial control task. Figure 3C plots the learning curves for the category- and the paired-associate-training group in all three tasks, in blocks of eight trials. As noted, the control task was meant to serve as a rule-discovery task; therefore, data from this task were not analyzed further.

Table 1

Participants' Accuracy (Percent Correct) in the three Type II Tasks

	Control task		Label task		Ideogram task	
Training Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
Category	88.8	7.2	97.4	2.6	97.9	2.4
Paired-associate	85.3	8.5	94.9	4.9	94.6	6.7

In analyzing categorization accuracy we created a dummy variable (“igt”) coding the interaction of training group by task (label vs. ideogram). The resulting factor had four levels: category-label, category-ideogram, paired-label, and paired-ideogram. Trial was normalized as in Experiment 1. Model comparison procedures suggested that the most parsimonious model was the one with a single smooth term, irrespectively of the level of the interaction factor, and with individual variation modeled by by-participant random smooth terms of trial. In particular, inclusion of separate smooth terms of trial (by igt) led to an increase in both fREML score (by 226.9) and estimated degrees of freedom (by 6). Thus, in R notation, the best-fitting model was:

```
acc ~ 1 + igt + s(trial) +  s(trial, sbj, bs = "fs", m = 1)
```

with s(trial) denoting a smooth term of trial, and s(trial, sbj, bs = "fs", m = 1) denoting by-participant random smooth terms of trial.

Figure 3D plots the model's estimates for the difference in accuracy between the label and ideogram tasks for the category- and the paired-associate-training group. Participants in the category-training group were less accurate in the label compared to the ideogram task over the entire range of trials. No difference was observed following learning to associate: Participants in the paired-associate training group were equally accurate in solving the label and ideogram task.

Discussion

Experiment 2 revealed a label advantage during learning to categorize, similarly to Experiment 1. A label advantage—importantly, of similar magnitude—was also revealed during learning to form associations. These results are consistent with an account of a general facilitation in processing verbal compared to hard-to-name stimuli (e.g., Tolins and Colunga, 2015) and do not provide support to the label-feedback hypothesis positing a categorization-specific mechanism (Lupyan, 2012a; 2012b).

In contrast, with respect to the sustained effects of labels for categories on attention, there was evidence suggesting that shapes that had previously been diagnostic of named categories captured attention to a greater extent, compared to shapes that had previously been predictive of hard-to-name categories. Specifically, Phase 1 category training was followed by lower accuracy in the label compared to the ideogram task in Phase 2, in accordance with our prediction. This sustained effect of labels on attention was found to be specific to the learning of categories: Following learning to associate in Phase 1 there was no evidence of differential processing of the trained shapes in Phase 2. These results argue in favor of a categorization-specific mechanism affecting the processing of diagnostic features in the long term.

Experiment 3

Experiment 2 provided evidence suggesting that labels for the categories have sustained effects on attention mechanisms recruited during categorization test tasks. Given the novelty of our results and also previous reports of absence of such effects (Tolins & Colunga, 2015), we sought to further examine the finding of sustained effects of category labels on attention. We therefore designed a different post-training test task, which involved the simultaneous presentation of the label and ideogram shapes and the monitoring of participants' eye movements. In this way, we hoped to examine the capture of attention more directly and provide a methodologically independent verification of our hypothesis.

In Experiment 3, participants were first administered either category or paired-associate training in a session identical to that of Experiment 2 (see Figure 1C). As before, the mechanism of selective activation of diagnostic features due to labels was assumed to be specific to the learning of categories, therefore the label advantage during learning to categorize was predicted to exceed a purported label advantage during learning to associate.

Following training, participants performed a visual discrimination task on the trained shapes, while their eye movements were monitored (cf. Farrell, 1985; Belke & Meyer, 2002). Based on the findings of Rehder and Hoffman (2005a; 2005b), fixation durations were treated as measures of attention during learning. We further applied this rationale to posttraining tasks: Insofar as the shapes that had previously been predictive of named categories would capture attention to a greater extent, compared to the shapes that had previously been predictive of hard-to-name categories, participants should spend more time fixating the label shapes than the ideogram shapes in trials presenting one label and one ideogram shape. In contrast, attention was assumed to be equally captured by shapes that had previously been associatively paired with either named or hard-to-name response cues. Therefore, participants in the paired-associate-training group were predicted to spend comparable amounts of time fixating the label and ideogram shapes.

Method

In a first phase, participants in Experiment 3 were administered either category or paired-associate training. In a second phase immediately following the training session, they were asked to perform a visual discrimination task, employing the previously trained shapes (see Figure 1C), while their eye movements were monitored.

Materials, procedure, and data analysis of the training session were identical to those described in Experiment 2. Thus, only the method of the visual discrimination task will be described here.

Participants. Overall 69 students of the University of Athens participated in exchange for course credit, meeting the requirements of normal or corrected-to-normal vision, no diagnosis of dyslexia, and Greek being their native language. The data from 21 participants were discarded prior to any data analysis due to technical failures of the eye tracker or reduced accuracy caused by eye glasses or contact lenses. Thus, results reported here correspond to a sample of 48 students. Twenty four (four male) of them were randomly assigned to the categorization training group (age M = 20.3 years; SD = 1.5) and 24 (three male) to the paired-associate-training group (age M = 22.5 years, SD = 5.9).

Materials.

***Discrimination session*.** The association items used in the paired-associate training task were presented as stimuli for the visual discrimination task. All four stimuli subtended a rectangle of roughly 2.5 cm horizontally by 6 cm vertically (or 1.8° × 4.4° of visual angle), presented on a 20-inch flat LCD monitor with a 1600 × 900 resolution at 60 Hz. Stimuli were placed 13.6 cm (or 10° of visual angle) to the left and right of center, to minimize the effectiveness of peripheral vision (cf. Belke & Meyer, 2002) and encourage eye movements. Random jitter—both horizontally and vertically—of maximum ±20 pixels (0.4°of visual angle) for both stimuli introduced a slight uncertainty about exact position to prevent iconic-memory strategies from dominating performance and—again—encourage eye movements.

Procedure.

***Discrimination session*.** Participants were administered four blocks of 24 trials, programmed in Experiment Builder software (SR Research Ltd.). Each stimulus was presented equally often within a block. Half of the trials were *different*, that is, presented two different stimuli on the screen, and the other half were *same*, that is, presented the same stimulus on both locations. All possible permutations of stimuli were included within a block. In pilot testing participants were found not to fixate a stimulus if it had just been presented on the same side of the screen. Thus, the pseudorandom trial order was constrained to preclude presentation of the same stimulus on consecutive trials (following Belke & Meyer, 2002).

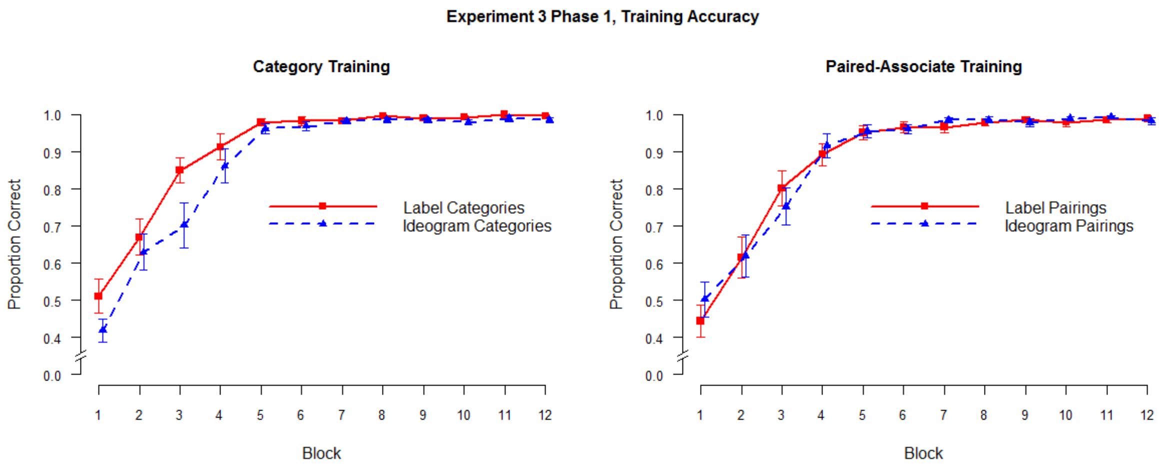
A discrimination trial started with a drift check, followed by a fixation cross—subtending a square with a side of 1° of visual angle—presented for a minimum of 500 ms. Presentation of the two shapes was triggered by the participants' gaze recorded within the fixation cross for 150 ms, and lasted for 3000 ms. Participants were required to press one of two keys on the keyboard—as fast and accurately as possible—to denote whether the two shapes were different or the same.

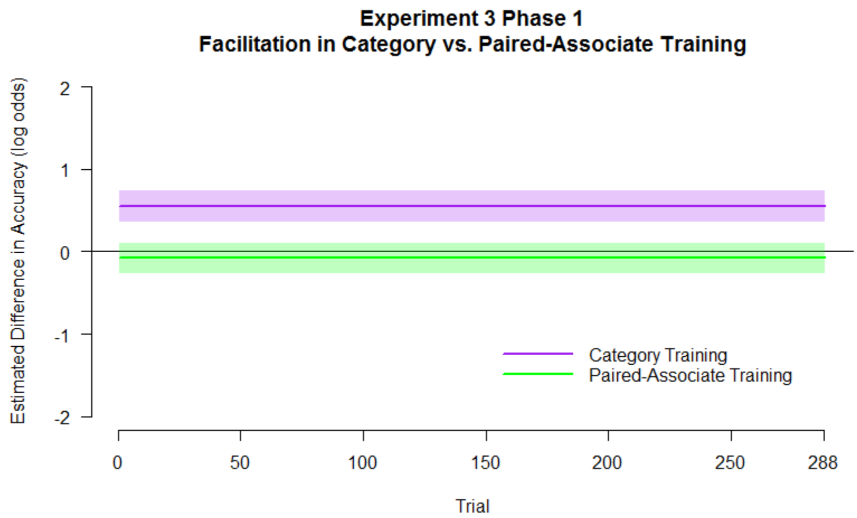
An Eyelink 1000 Plus eyetracker sampling monocularly at 2000 Hz recorded the eye providing the best calibration accuracy. A head and chin rest was used, and calibration took place on average every two blocks, or more often if required. Participants were given a block of practice trials, there was a short break in the middle of the procedure, and the task lasted on average 20 minutes.

Data preprocessing and analysis.

Discrimination session. Analyses of participants' eye movements focused on fixation duration and only included data from trials presenting one label and one ideogram shape (8 trials within each block). The online parser of SR Research Ltd was used for fixation detection. Two rectangular areas of interest (AOIs) were defined prior to data collection, each subtending a square exceeding each stimulus by a margin of 5.5° of visual angle (7.5 cm). This margin was defined as the sum of the equipment's nominal accuracy (0.5° of visual angle) and a rough measure of the span of peripheral vision (5° of visual angle). Because of the substantial eccentricity of stimulus placement near the screen edges, the AOIs were not symmetric (only 100 pixels, amounting to 2.78 cm, or 2° of visual angle, for the distal margins). Duration of fixations within an AOI was determined using the Data Viewer software (SR Research Ltd.).

Following Henderson, Weeks, and Hollingworth (1999; Võ & Henderson, 2009), fixations with duration less than 90 ms or greater than 1000 ms were excluded from analysis. The sum of durations of fixations landing within an AOI (hereafter "dwell time") was calculated for each participant and trial. Next, average dwell time by block was calculated by averaging fixation duration over trials within a block for each participant and block, aiming to account for temporal order effects (such as increasing familiarity; cf. Lupyan & Spivey, 2008, Supplementary Materials). Average dwell time was analyzed using the ez package (Lawrence, 2015) for ANOVA and the multcomp package (Hothorn, Bretz, & Westfall, 2008) for post-hoc multiple comparisons. All analyses were conducted in R version 3.3.3 (R Core Team, 2017).


Experiment 3 Phase 1, Training Accuracy
Category Training
Paired-Associate Training
Proportion Correct
Label Categories
Ideogram Categories
Label Pairings
Ideogram Pairings
Block


Experiment 3 Phase 1
Facilitation in Category vs. Paired-Associate Training
Estimated Difference in Accuracy (log odds)
Category Training
Paired-Associate Training
Trial

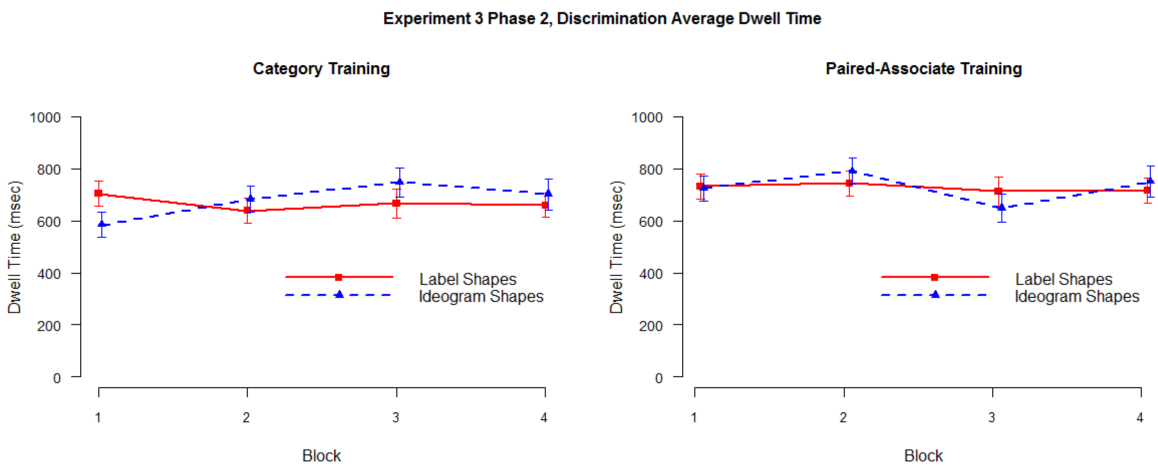

Experiment 3 Phase 2, Discrimination Average Dwell Time
Category Training
Paired-Associate Training
Dwell Time (msec)
Label Shapes
Ideogram Shapes
Block

Figure 4. Results of Experiment 3. (A) Training accuracy of the label and ideogram categories (left panel) and of the label and ideogram pairings (right panel) in blocks of 24 trials. Error bars show between-subjects standard errors of the means. (B) Estimated difference in accuracy (in log odds) between the label and ideogram categories and the label and ideogram pairings, excluding random effects of participants. Error bands show 95% confidence intervals of the estimates. (C) Discrimination task, average dwell time in the labels and ideogram shapes for the category-training group (left panel) and the paired-associate-training group (right panel) in blocks of eight trials. Error bars show between-subjects standard errors of the means.

Results

Training session. No-response trials (3 from the category-training group and 4 from the paired-associate-training group, comprising 0.05% of the data) and trials with response latencies less than 250 ms (5, from paired-associate-training only; 0.03%) were excluded from analyses. Participants' accuracy increased as trials progressed in learning both the label and ideogram categories, averaging 88.8% (*SD* = 5.8) correct in category learning and 88.3% (*SD* = 5.6) in paired-associate learning (label categories: *M* = 90.5%, *SD* = 5.7; ideogram categories: *M* = 87.1%, *SD* = 6.8; label pairings: *M* = 87.9%, *SD* = 6.5; ideogram pairings: *M* = 88.4%, *SD* = 6.6). Figure 4A depicts participants' learning curves for the category- and the paired-associate-training group in blocks of 24 trials.

As in Experiment 2, accuracy was analyzed using generalized additive mixed models with a normalized trial predictor and a dummy variable (“igr”) coding the interaction of group by response cue. The most parsimonious model included a single smooth term of trial, accounting for all training group and response cue combinations, in addition to by-participant random smooth terms of trial. In particular, inclusion of separate smooth terms of trial for each level of the igr factor led to an increase in both the fREML score (by 108.1) and the estimated degrees of freedom (by 6). In R notation, the best-fit model was specified as:

```
acc ~ igr + s(trial) + s(trial, sbj, bs = "fs", m=1)
```

with s(trial) denoting a smooth term of trial, and s(trial, sbj, bs = "fs", m=1) denoting by-participant random smooth terms of trial.

Figure 4B plots the model's estimate of the facilitation due to labels in the category- and the paired-associate-training group, in log odds, excluding random effects of participants. Labels facilitated category learning throughout the training procedure. There was no label facilitation in the paired-associate group: Participants were equally accurate in learning the label and ideogram pairings throughout the task.

Discrimination session.

Accuracy. No-response trials (11 from the category-training group and 5 from the paired-associate-training group, totaling 0.34% of the data) were excluded from analysis. Both groups were highly accurate in the task: Taking all trials into account, participants averaged 98.1 % (*SD* = 1.9) correct responses in the category-training group and 98.3 % (*SD* = 1.6) in the paired-associate group. A Welch's t-test for independent samples suggested that the two groups did not differ in accuracy, $t(44.08) = -.28$, $p = .783$. To examine the differences between *same* and *different* trials we conducted a two-way ANOVA on average accuracy, with group as the between-subjects factor, and type of trial (*same* vs. *different*) as the within-subjects factor. Results revealed no interaction of group by type, $F(1, 46) = 1.392$, $\eta^2 = .014$, $p = .244$, and no effect of group, $F(1, 46) = .075$, $\eta^2 = .0008$, $p = .785$. There was an effect of trial type, $F(1, 46) = 6.769$, $\eta^2 = .066$, $p = .013$, suggesting higher accuracy in *different* than in *same* trials, consistent with previous findings (Farell, 1985). Further analyses revealed no effect of response cues on either accuracy or response latency (analyses available upon request).

Average dwell time. This analysis included only critical trials, that is, “different” trials displaying one label shape and one ideogram shape (8 per block). Participants in the category-training group spent on average 668.28 ms (*SD* = 213.09) fixating the label shapes and 683.82 ms (*SD* = 220.49) fixating the ideogram shapes. Participants in the paired-associate-training group spent on average 726.59 ms (*SD* = 225.21) fixating the labels shapes and 727.52 ms (*SD* = 205.96) fixating the ideogram shapes.

Figure 4C plots average dwell time on the label and ideogram shapes, for the category- and the paired-associate-training group, per block of trials. A three-way repeated measures ANOVA was conducted on dwell time with group as the between-subjects factor, and shape (label vs. ideogram) and block (1, 2, 3, 4) as within-subject factors. There was a three-way interaction, $F(3,138) = 3.611$, $\eta^2 = .008$, $p = .002$, therefore we analyzed data from each group separately. In a two-way ANOVA for the category-training group with shape and block as within-subjects factors there was no effect of shape, $F(1, 23) = 0.311$, $\eta^2 = .0006$, $p = .582$, or of block of trials, $F(3, 69) = 1.055$, $\eta^2 = .009$, $p = 0.374$, but—importantly—there was an interaction of shape by block, $F(3, 96) = 7.269$, $\eta^2 = .024$, $p = .0003$. Multiple comparisons per block—with adjusted p values—showed that participants spent more time fixating the label ($M = 704.31$, $SD = 236.99$) compared to the ideogram shapes ($M = 584.57$, $SD = 235.75$) for the first block of trials, $\beta = 118.27$, $z = 2.684$, $p = .025$. There was no difference in average dwell time for subsequent blocks (second: $\beta = -45.58$, $z = -0.825$, $p = .807$; third: $\beta = -83.12$, $z = -1.505$, $p = .348$; fourth: $\beta = -44.26$, $z = -1.004$, $p = .6857$). A similar analysis for the paired-associate training group revealed no interaction of shape by block, $F(3, 69) = 1.287$, $\eta^2 = .008$, $p = .286$, and no main effect of either block, $F(3, 69) = 2.500$, $\eta^2 = .015$, $p = .067$, or—importantly—shape, $F(1, 23) = 0.005$, $\eta^2 = .00002$, $p = .942$.

Discussion

Similarly to Experiments 1 and 2, a label advantage was found during learning to categorize. Contrary to Experiment 2, the label advantage was found to be specific to the learning of categories: There was no facilitation due to labels during learning to associate. Overall, these results are supportive of the assumption of a categorization-specific mechanism offering facilitation due to labels (Lupyan 2012a; 2012b). The discrepancy between training results across Experiments 2 and 3 is taken up in more detail in the General Discussion.

Experiment 3 also examined the sustained effect of category labels on attention, contrasted with the sustained effects of labels for associations. Results suggested that early in the task (in the first block of trials) shapes that had previously been predictive of named categories captured attention to a greater extent compared to shapes that had previously been predictive of hard-to-name categories. No effect was found later in the task, suggesting that the sustained effects of labels may be subject to participants' adaptation to the task, an issue considered further in the General Discussion. Importantly, the sustained effects of labels were specific to the learning of categories: Attention was equally captured by shapes that had previously been paired with either labels or hard-to-name symbols in paired-associate training. Therefore, the results of Experiment 3—similarly to the results of Experiment 2 —are consistent with the hypothesis of sustained effects of category labels, contrary to previous research (Tolins & Colunga, 2015).

General Discussion

In a series of three experiments, we tested the label-feedback hypothesis using a within-subjects design to examine whether the label advantage and its sustained effects are specific to categorization or apply to associations as well.

Label Advantage During Learning

A label advantage was consistently observed in learning to categorize, over all 3 experiments, which cannot be attributed to special selection of experimental material (cf. Lupyan & Casasanto, 2015), given our counterbalanced materials and procedures. A label advantage was also observed in paired-associate learning in Experiment 2 but not in Experiment 3.

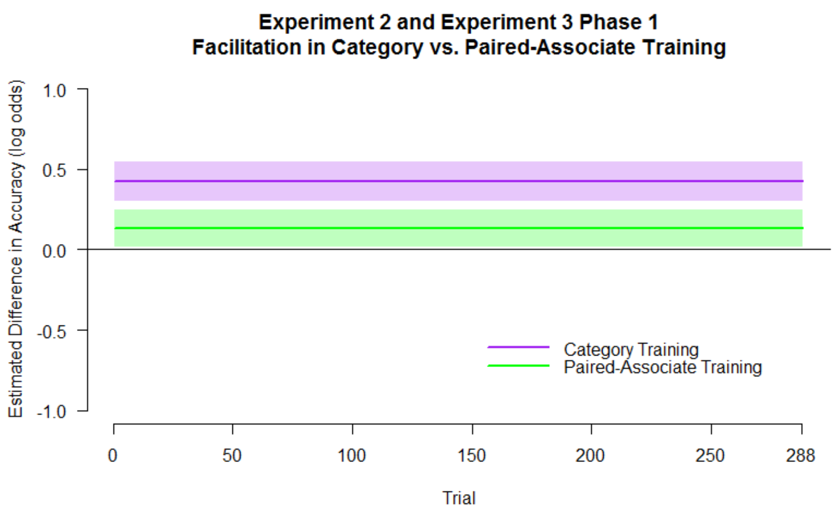


Figure 5. Analysis of pooled data from the training sessions of Experiment 2 and Experiment 3. Estimated difference in accuracy (in log odds) between the label and ideogram categories and the label and ideogram pairings, excluding random effects of participants. Error bands show 95% confidence intervals of the estimates.

Pooling the Phase 1 data from Experiments 2 and 3 (see Figure 5) suggests that the label advantage was greater during learning to categorize than during learning to associate (detailed analyses available from the authors). This suggests that the mechanism of selective activation of diagnostic features (Lupyan, 2012a; 2012b) may be complementary to a more general account of verbal labels serving as material symbols (Brojde et al., 2011; Lupyan et al., 2007; Tolins and Colunga, 2015). That is, labels may facilitate cognitive processing in general, as suggested by the observed label advantage during associative learning. In addition, selective activation provides further facilitation during learning to categorize. Whether the label advantage during learning to associate and categorize stem from overlapping or entirely distinct mechanisms, our results are consistent with a categorization-specific mechanism, thus supporting the label-feedback hypothesis (Lupyan, 2012a; 2012b).

Long-term Effects Of Labels On Attention

To examine the long-term effects of category labels on attention, participants in Experiment 2 were administered Type II (Shepard et al., 1961) categorization tasks using the shapes previously associated with, or categorized by, label and ideogram response cues. Results from the category-training group showed that participants were less accurate in the label compared to ideogram task, consistent with the assumption of long-term effects of category labels on attention. Importantly, these effects were found to be categorization-specific: Participants administered paired-associate learning were subsequently equally accurate in the label and ideogram tasks. Results from Experiment 2 are—to the best of our knowledge—the first evidence suggesting long-term effects of category labels (but not of labels for the associations) on attention mechanisms.

In Experiment 3, we contrasted the sustained effect of labels for the categories with those for the associations by measuring participants' fixation durations on shapes that had previously been predictive of named categories compared to shapes that had been predictive of hard-to-name categories. The results showed that, following learning to categorize, participants spent greater time fixating the label compared to the ideogram shapes for the first block of the visual discrimination task. This difference disappeared in subsequent blocks, suggesting that the effect of category labels on attention was transient and did not survive the accumulation of experience within the discrimination task. This comes as no surprise given previous explorations of the idea that visual processing—

affected by labels of overlearned categories—might depend on experimental trial (Lupyan & Spivey, 2008), or the finding that the effect of overtly presenting labels of categories is time-dependent (Lupyan & Spivey, 2010b). Moreover, analysis of response latencies in the discrimination task revealed that participants' speed of responding increased as trials progressed. Given the tight coupling of behavioral and eye-movement measures (Rehder & Hoffman, 2005a; 2006b) it seems plausible to also expect practice effects in fixation durations. In contrast, participants in the paired-associate group spent equal time fixating the shapes that had previously been paired with names compared to stimuli that had previously been paired with hard-to-name symbols. Thus, similarly to Experiment 2, the sustained effect of labels on attention were only found after learning to categorize, not after learning to associate.

We take the findings of Experiments 2 and 3 to be supportive of the label-feedback hypothesis, positing a categorization-specific mechanism inducing sustained effects on attention processes recruited during visual processing of the diagnostic features. However, previous exploration of long-term effects of category labels by Tolins and Colunga (2015) revealed that attention processes were not affected by redundant labels. A possible reconciliation of our findings with those results may be based on the degree of learning the labels. Specifically, Lupyan (2006) suggested that it is the efficient learning of labels—rather than their mere presence—that affects learning (see also Brojde & Colunga, 2011 for results supporting this argument). Thus, if participants in the Tolins and Colunga study did not learn the labels sufficiently well, it is to be expected that neither initial effects on categorization accuracy nor sustained effects on attention were revealed. In contrast, in our experiments labels were not redundant; rather, they were the only available response cues and could not have been ignored by participants. Thus, both initial and sustained effects were revealed. We submit that further investigation of the label-feedback hypothesis should include an assessment of the degree to which participants have learned the labels prior to examination of label effects on learning.

Dimensions versus Features in Learning to Categorize

An important implication of the present results concerns the distinction between perceptual dimensions and perceptual features in learning to categorize. The majority of experimental research examining category-learning processes and systems has utilized between-subjects manipulations and corresponding comparisons (e.g., Ashby & Maddox, 2005). Although this approach has proven fruitful in advancing our understanding of category learning, it does not help elucidate whether it is entire perceptual dimensions or, rather, specific perceptual features that are important for categorization. A perceptual dimension encompasses all possible values within it; therefore, even if it is the features of specific values that capture attention during learning to categorize, a between-subjects design is—in principle—not diagnostic of the distinction and can only attest in favor of dimensional sensitization or activation.

Surprisingly few studies have addressed the dimension vs. features distinction. As noted, Goldstone (1994) showed that perceptual sensitization following learning to categorize is a localized phenomenon (i.e., it is greater for values of a diagnostic dimension that cross a category boundary compared to values that belong to the same category), and this result was replicated by Van Gulick and Gauthier (2014). In related research, Aha and Goldstone (1992) provided evidence suggesting that, following learning to categorize, different values of a perceptual dimension may be selectively attended to (see also Blair, Watson, Walshe, & Maj, 2009).

With respect to the label-feedback hypothesis, Lupyan (2012b) explicitly posited that it is specific perceptual features that are selectively activated by verbal labels, rather than general perceptual dimensions. However, the studies examining the initial and sustained effects of category labels (Brodje et al., 2011; Lupyan & Casasanto, 2015; Lupyan et al., 2007; Perry & Lupyan, 2014; Tolins & Colunga, 2015) have all used between-subjects manipulations and—naturally—concluded that labels result in the increased capturing of attention by perceptual dimensions. In contrast, in our experiments the varying nameabilty of formed categories was a within-subjects manipulation. All participants learned both named and hard-to-name categories in a single training procedure. If labels activate the diagnostic perceptual dimension as a whole, rather than the features of specific diagnostic values linked to labels (i.e., the dimension of shape rather than the label shapes specifically), then we should have observed no difference in accuracy between label and ideogram category training, as well as no difference in the posttraining processing of label and ideogram shapes, in any of our experiments. We may therefore conclude that labels for the categories result in increased capturing of attention (both during and following learning) by diagnostic perceptual *features,* rather than dimensions, in accordance with Goldstone's

(1994) suggestion that perceptual space is locally warped as a result of learning to categorize. We also suggest that within-subjects manipulations can be a useful methodological approach in allowing the examination of hypotheses that are not falsifiable in typical between-subjects designs.

Limitations, Concerns, And Directions For Future Research

In examining the label-feedback hypothesis (Lupyan, 2012a; 2012b) we manipulated linguistic activity by using names vs. hard-to-name symbols, rather than by using redundant labels vs. the absence of labels (e.g., Brojde et al., 2011; Lupyan et al., 2007; Tolins & Colunga, 2015). This manipulation took the effect of correlated cues out of the equation (cf. Lupyan et al.) but introduced a possible limitation. Verbal labels and ideograms were equated in size but arguably placed different demands on, e.g., memory or perception, potentially leaving the results open to alternative interpretations. Similar asymmetries are seen in previous studies (for example, between geometric and resistant-to-verbalization stimulus features, Kurtz et al., 2013, or between verbal labels and location cues, Lupyan et al., 2017), as it is not always clear what should be equated and by which criteria. Further theoretical and experimental work should address criteria and procedures for equating verbal and hard-to-name stimuli.

A second limitation of our study might arise from using shape as the category-diagnostic dimension, as the effect of category labels can be moderated by the choice of dimension (Brojde et al., 2011). Additionally, it has been shown that labels shift attention to dimensions that are historically predictive of category membership (e.g., to shape rather than hue, Brodje et al., or to frequency rather than orientation, Perry & Lupyan, 2014, see also Perry & Lupyan, 2016, for related research). Thus, it is possible that the effects of labels are specific to the perceptual dimensions that are category-diagnostic in real-world situations. Participants in our experiments also relied on shape (which has a special status for categorization; e.g., Landau, Smith, & Jones, 1988) to learn categories or associations. A more stringent test of the label-feedback hypothesis would entail examining the initial and sustained effects of category labels when learning is based on historically non-predictive dimensions, such as, e.g., color or orientation.

An intriguing discrepancy between Experiment 2 and 3 concerns the temporal duration of the observed sustained effects of category labels on attention mechanisms. In

Experiment 2 participants administered category training were subsequently found to be less accurate in the label compared to the ideogram task for the entire range of trials. In contrast, the category-training group of Experiment 3 spent more time fixating the label compared to the ideogram shapes only at the beginning of the discrimination task (for only the first out of totally four blocks of trials). Learning to solve the Type II task and performing a visual discrimination task arguably pose different processing demands, therefore observing a different temporal pattern of an effect across these tasks does not seem troublesome, at least in this preliminary research. Nevertheless, future studies should examine in more detail the factors influencing the longevity of the sustained effects of labels on attention.

Our results suggest that verbal labels for the categories, but not for associations, have long-term effects on attention processes. This effect was predicted on the basis of converging evidence from perceptual learning (Folstein et al., 2015; Goldstone, 1994), learned attention (Goldstone & Steyvers, 2001; Krushke, 1996), and the label-feedback hypothesis (Lupyan, 2012a; 2012b). Our experiments were not designed to distinguish between these accounts, and thus the origin of the effect remains unclear. We argue that the perceptual-learning and the learned-attention accounts are complementary, in that a cognitive system (be it perceptual or attentional) is shaped during learning and keeps on exerting an influence following learning. Attesting to this assumption, in the study of category learning, attention and perception are not clearly distinguished, so what is called “learned attention” may suggest “perceptual warping” and vice versa. For example, Goldstone (1994) interpreted findings of perceptual warping as indicating competition for attention, and Folstein et al. (2015) considered findings of learned attention as suggesting perceptual modulation.

There is, however, a critical difference between these two accounts—namely, perceptual learning and learned attention—and the label-feedback hypothesis, in that according to the latter hypothesis the sustained effects of labels are attributed to the self-activation of learned labels (Lupyan, 2012a). Suppressing participants' linguistic activity during a posttraining task (see Perry & Lupyan, 2013, for manipulations other than the commonly used verbal interference) might elucidate the origin of the observed sustained effects. If the effects survive the down-regulation of participants' linguistic activity, they may be attributed to either the perceptual learning or the learned attention account. In the opposite case, the sustained effects of category labels may be attributed to the labels-dependent mechanism of selective activation of diagnostic features, supporting the label-feedback hypothesis, an issue left open for future research.

Finally, an intriguing result of Experiment 2 concerns the emergence of differences in posttraining categorization despite similar accuracy during initial learning. A two-way ANOVA on categorization accuracy in the three Type II tasks, with group (category vs. paired-associate training) as the between-subjects factor, and task (control, label, ideogram) as the within-subjects factor, revealed no interaction of group by task, $F(2, 124) = 0.348$, $\eta^2 = .003$, $p = .707$. There was a main effect of task, $F(2, 124) = 68.329$, $\eta^2 = .384$, $p < .001$, reflecting participants' greater difficulty with “solving” the first compared to the following two tasks. Interestingly, there was also a main effect of group, $F(1, 62) = 4.911$, $\eta^2 = .037$, $p = .031$, suggesting that the category-training group was more accurate than the paired-associate-training group in all three Type II tasks. This cannot be accounted for by group-level differences in general learning capacities, because a oneway ANOVA on training accuracy revealed no effect of group, $F(1,62) = 0.701$, $\eta^2 = .011$, $p = .406$. No other posttraining group differences were found in any analyses of Experiment 3. Perhaps the discrimination task was too easy (recall that participants were all highly accurate in discriminating stimuli) for an effect of training regime (category vs. paired-associate) to manifest itself. It remains to be seen if different effects on posttraining performance may emerge in more demanding learning tasks.

Conclusions

In the present study we sought to test the label-feedback hypothesis (Lupyan, 2012a; 2012b) by investigating the effect of labels for the categories during initial learning and in posttraining test tasks. In a series of three experiments, we found that participants were more accurate in learning named compared to hard-to-name categories. It was also revealed that the label advantage during learning to categorize was of greater magnitude, compared to a label advantage during learning to associate, precluding explanation by a general theory of facilitation in processing verbal stimuli. In contrast to previous research (Tolins & Colunga, 2015), there was evidence of sustained effects of category labels on attention mechanisms recruited in two test tasks: a modified version of the Type II (Shepard et al. 1961) task, and an eyetracking visual discrimination task. These sustained effects were specific to categorization, in that no effects emerged following learning to associate.

The present research contributes to the category-learning literature by suggesting that linguistic representations are not isolated entities restricted to encapsulated language modules (e.g., Fodor, 1985) but, rather, interact with perceptual or attention processes recruited both during and also after learning to categorize (Lupyan, 2012a; 2012b). Our results have challenging implications for current theories of learning, and also for the language and thought debate (e.g., Gleitman & Papafragou, 2013; Regier, Kay, Gilbert, & Ivry, 2010), helping elucidate the more general question regarding the interplay between the language faculty and learning processes.

References

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 534–539). Hillsdale, NJ: Erlbaum

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology, 56*, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Baayen, R. H. (2013). Multivariate statistics. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 337–372). New York, NY: Cambridge University Press.

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"–"different" decisions. *European Journal of Cognitive Psychology, 14*(2), 237–266. http://dx.doi.org/10.1080/09541440143000050

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1196–1206. http://dx.doi.org/10.1037/a0016272

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience, 35*(25), 9329–9335. http://dx.doi.org/10.1523/JNEUROSCI.5111-14.2015

Brojde, C. L., Porter, C., & Colunga, E. (2011). Words can slow down category learning. *Psychonomic Bulletin & Review, 18*(4), 798–804, http://dx.doi.org/10.3758/s13423-011-0103-z

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447–456. http://dx.doi.org/10.1016/j.jml.2007.11.004

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition, 143*, 93–100. http://dx.doi.org/10.1016/j.cognition.2015.06.008

Farell, B. (1985). "Same"–"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin, 98*(3), 419–456. http://dx.doi.org/10.1037/0033-2909.98.3.419

Fodor, J. A. (1985). Précis of the modularity of mind. *Behavioral and Brain Sciences, 8*(1), 1–42. http://dx.doi.org/10.1017/S0140525X0001921X

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object representations: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 807–820. http://dx.doi.org/10.1037/a0025836

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex, 23*(4), 814–823. http://dx.doi.org/10.1093/cercor/bhs067

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2014). Perceptual advantage for category-relevant perceptual dimensions: the case of shape and motion. *Frontiers in Psychology, 5*, 1394. http://dx.doi.org/10.3389/fpsyg.2014.01394

Folstein, J. R., Palmeri, T. J., Van Gulick, A. E., & Gauthier, I. (2015). Category learning stretches neural representations in visual cortex. *Current Directions in Psychological Science, 24*(1), 17–23. http://dx.doi.org/10.1177/0963721414550707

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124. http://dx.doi.org/10.3758/BF03195503

Fotiadis, F. A., & Protopapas, A. (2014). The effect of newly trained verbal and nonverbal labels for the cues in probabilistic category learning. *Memory & Cognition, 42*(1), 112–125. http://dx.doi.org/10.3758/s13421-013-0350-5

Fotiadis, F. A., & Protopapas, A. (2016). No effect of verbal labels for the shapes on Type II categorization tasks. In Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J.C. (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 277–282). Austin, TX: Cognitive Science Society.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition, 105*(1), 218–228. http://dx.doi.org/10.1016/j.cognition.2006.09.005

Gleitman, L. R. & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *Handbook of Cognitive Psychology* (pp. 504–523). New York, NY: Oxford University Press.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178–200. http://dx.doi.org/10.1037/0096-3445.123.2.178

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*(1), 116–139. http://dx.doi.org/10.1037/0096-3445.130.1.116

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228. http://dx.doi.org/10.1037/0096-1523.25.1.210

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. http://dx.doi.org/10.1002/bimj.200810425

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*(2), 150–166. http://dx.doi.org/10.1016/j.jecp.2006.09.002

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

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22–44. http://dx.doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*(2), 225–248. http://dx.doi.org/10.1080/095400996116893

Kurtz, K. J., Levering, K., Romero, J., Stanton, R. D., & Morris, S. N. (2013). Human learning of elemental category structures: Revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 552–572. http://dx.doi.org/10.1037/a0029178

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*(3), 299–321. http://dx.doi.org/10.1016/0885-2014(88)90014-7

Lawrence, M. A. (2015). *ez: Easy Analysis and Visualization of Factorial Experiments*. R package version 4.3. Retrieved from https://CRAN.R-project.org/package=ez

Levine, M. (1975) *A cognitive theory of learning: Research on hypothesis testing*. Hillsdale, NJ: Erlbaum.

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition, 31*(5), 790–799. http://dx.doi.org/10.3758/BF03196117

Lupyan, G. (2006). Labels facilitate learning of novel categories. In A. Cangelosi, A.D.M. Smith, & K. Smith (Eds.), *The Sixth International Conference on the Evolution of Language* (pp. 190–197). Singapore: World Scientific.

Lupyan, G. (2008a). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137*(2), 348–369. http://dx.doi.org/10.1037/0096-3445.137.2.348

Lupyan, G. (2008b). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*(2), 566–577. http://dx.doi.org/10.1016/j.cognition.2008.03.009

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review, 16*(4), 711–718. http://dx.doi.org/10.3758/PBR.16.4.711

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition, 3*, 54. http://dx.doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Waltham, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394293-7.00007-8

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition, 7*(2), 167-193. http://dx.doi.org/10.1017/langcog.2014.21

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex, 49*(5), 1187–1194. http://dx.doi.org/10.1016/j.cortex.2012.06.006

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking redundant labels facilitate learning of novel categories. *Psychological Science, 18*(12), 1077–1083. http://dx.doi.org/10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology, 18*(10), R410–R412. http://dx.doi.org/10.1016/j.cub.2008.02.073

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Auditory cues facilitate visual object detection. *PLoS ONE, 5*(7), e11452. http://dx.doi.org/10.1371/journal.pone.0011452.

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72*(8), 2236–2253. http://dx.doi.org/10.3758/APP.72.8.2236.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General, 141*(1), 170–186. http://dx.doi.org/10.1037/a0024904

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing*. Psychological Science, 21*(5), 682-691. http://dx.doi.org/10.1177/0956797610366099

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 14196-14201. http://dx.doi.org/10.1073/pnas.1303312110

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126–135. http://dx.doi.org/10.1037/0278-7393.14.1.126

Mathy, F., Haladjian, H. H., Laurent, E., & Goldstone, R. L. (2013). Similarity-dissimilarity competition in disjunctive classification tasks. *Frontiers in Psychology, 4*(26), 1–14. http://dx.doi.org/10.3389/fpsyg.2013.00026

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and nonrule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1518–1533. http://dx.doi.org/10.1037/a0013355

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 117–162). San Diego: Academic Press. http://dx.doi.org/10.1016/S0079-7421(10)52003-6

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. Journal *of experimental psychology: General, 115*(1), 39–57. http://dx.doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22*(3), 352–369. http://dx.doi.org/10.3758/BF03200862

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral dimension stimuli. *Psychonomic Bulletin & Review, 3*(2), 222–226. http://dx.doi.org/10.3758/BF03212422

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13, http://dx.doi.org/10.1016/j.jneumeth.2006.11.017

Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience, 7*, 122. http://dx.doi.org/10.3389/fnbeh.2013.00122

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language, 135*, 66–72. http://dx.doi.org/10.1016/j.bandl.2014.05.005

Perry, L. K., & Lupyan, G. (2016). Recognising a zebra from its stripes and the stripes from "zebra": the role of verbal labels in selecting category relevant information. *Language, Cognition and Neuroscience*. Advanced online publication. http://dx.doi.org/10.1080/23273798.2016.1154974

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso, J., Myers, C. E., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*(6863), 546–550. http://dx.doi.org/10.1038/35107080

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources and Evaluation, 46*(3), 449–459. http://dx.doi.org/10.1007/s10579-010-9130-z

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Regier , T , Kay , P. , Gilbert , A. L. , & Ivry , R. B. (2010). Language and thought: Which side are you on, anyway? In B. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 165–182) . Oxford, England : Oxford University Press

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*(1), 1–41. http://dx.doi.org/10.1016/j.cogpsych.2004.11.001

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 811–829. http://dx.doi.org/10.1037/0278-7393.31.5.811

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42. http://dx.doi.org/10.1037/h0093825

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in Rhesus Monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General, 133*(3), 398–414. http://dx.doi.org/10.1037/0096-3445.133.3.398

Tolins, J., & Colunga, E. (2015). How words anchor categorization: conceptual flexibility with labeled and unlabeled categories‡. *Language and Cognition, 7*(2), 219–238. http://dx.doi.org/10.1017/langcog.2014.26

Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology, 57*(3), 147–154. http://dx.doi.org/10.1037/h0048723

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*(4), 584–589. http://dx.doi.org/10.3758/BF03193889

Van Gulick, A. E., & Gauthier, I. (2014). The perceptual effects of learning object categories that predict perceptual goals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1307–1320. http://dx.doi.org/10.1037/a0036822

Van Rij, J., Wieling, M., Baayen, R., and van Rijn, H. (2016). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs (R package version 2.2). Retrieved from https://cran.r-project.org/web/packages/itsadug/index.html

Võ, M. L. H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*(3), 24–24. http://dx.doi.org/10.1167/9.3.24.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology, 29*(3), 257–302. http://dx.doi.org/10.1006/cogp.1995.1016

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3–36. http://dx.doi.org/10.1111/j.1467-9868.2010.00749.x

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971–979. http://dx.doi.org/10.3758/PBR.15.5.971

Yoshida, H., & Smith, L.B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16*, 90–95. http://dx.doi.org/10.1111/j.0956-7976.2005.00787.x

5

General Discussion

In the present dissertation I sought to investigate the interplay between the language faculty and category-learning processes. In particular, I investigated the effect of verbal labels for (a) the to-be-categorized material, and (b) the formed categories. To this end, theories and findings from two different subfields within the study of category learning were combined, namely the dual-systems theories (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Poldrack & Foerde, 2008) and the perceptual learning account (Goldstone, 1994; Lupyan, 2012a; Lupyan, Rakison, & McClelland, 2007). In a series of experiments—detailed in Chapters 2,3, and 4—I assessed the effect of verbal labels on learning to categorize through novel training procedures that allowed me to contrast labels with hard-to-name visual symbols. Finally, in examining the effect of verbal labels I employed both typical response-latency studies and also eye tracking methodology.

The main findings of the experiments presented here may be summarized as follows: (a) Verbal labels for the auditory cues of a probabilistic categorization task were found to facilitate explicit hypothesis-testing processes of category learning (Chapter 2). In contrast, verbal labels for the features of categorization items in a visual deterministic category learning task were found to leave categorization processes unaffected (Chapter 3). This discrepancy is discussed in the next section. (b) Verbal labels for the formed categories were found to facilitate learning, and also affect subsequent processing of the category-diagnostic perceptual features (Chapter 4). Thus, the present dissertation may be said to support the notion that language interacts with category-learning processes, as evident both during the learning tasks and also in subsequent test tasks.

Limitations, Open Questions, and Directions for Future Research

control condition). In particular, the control group learned to associate shapes to verbal labels and to hard-to-name cues, in a within-subjects design. Results suggested that there was no difference in categorization accuracy between the Type II task utilizing the shapes associated with names and the task utilizing the shapes associated with hard-to-name symbols. Thus, the finding that names for the features of the categorized items did not affect learning during the solving of Type II category structure was consistently obtained across two experiments, one employing a between- (Chapter 3) and one a within-subjects design (Chapter4). To conclude, names for the categorized items were found to facilitate explicit category-learning processes, in Chapter 2, and to leave categorization performance unaffected, in Chapters 3 and 4.

It could be argued that the finding of hypothesis-testing facilitation due to names for the stimuli, as described in Chapter 2, was a false-positive result. This possibility may only be addressed empirically, so future research should examine the replicability of the finding of hypothesis-testing facilitation due to names for the auditory cues of the WPT.

Another possibility is that the discrepancy of results across Chapter 2 and Chapters 3 and 4 may be attributed to the differences across the studies. One important difference concerns the formal structure of the learning paradigms used to examine explicit learning processes. The WPT (Chapter 2) is characterized by a probabilistic structure, i.e., a combination of cues may predict different outcomes on different trials. In contrast, the Type II task (Chapters 3 and 4) has a deterministic structure, i.e., a categorization item consistently belongs to a specific category throughout the task. Research has not suggested that tasks characterized by probabilistic vs. deterministic structure recruit fundamentally different learning processes or systems. For example, the same manipulations have been utilized across probabilistic and deterministic structures to modulate learning system involvement (Foerde, Poldrack, & Knolwotn, 2007, and Zeithamova & Maddox, 2006). Nevertheless, an explanation on the basis of formal category structure cannot be precluded. Further empirical investigation should assess the possibility that the effect of names for categorization items (or their features) is structure-dependent. Another important difference across the studies of Chapter 2 and Chapters 3 and 4 concerns the learning modality. I utilized auditory cues in the WPT, and visual categorization items in the Type II tasks. It could be that names for the categorized items facilitate explicit learning processes in the auditory, but not in the visual modality. Indeed, research investigating the cognitive processes of statistical learning attests to this assumption by revealing differences in learning processes across modalities (Conway & Christiansen, 2009; Saffran, 2002). Further empirical investigation should examine the possibility that the effect of names for the stimuli on learning to categorize is modality-dependent.

Alternatively, the fact that there was no evidence suggesting rule-discovery facilitation due to names during the solving of Type II tasks may be attributed to a representational mismatch. As already noted in Chapter 3, informal reports suggested that during the solving of the Type II task participants based their categorization judgments on specific features of categorization items (i.e., the stimuli's corners) rather than on the shapes whose nameability had been manipulated. It stands to reason that if the representations of shapes were immaterial in solving the Type II tasks then names for the shapes will have no effect on category learning processes.

Finally, the discrepancy of results across the two training paradigms (the WPT and the Type II task) may be attributed to the nature of category learning processes recruited during the solving of the Type II tasks. Dual systems theorists have assumed that learning in the canonical version of the Type II task (i.e., the version employing highly familiar geometric stimuli) is mediated by explicit processes or rule discovery (e.g., Minda, Desroches, & Church, 2008; Minda & Miles, 2010; Smith, Minda, & Washburn, 2004). It has also been assumed that learning in the Type II tasks is mediated by similarity-based learning processes (Krushke, 1992; Nosofsky, Palmeri, McKinley, & Glauthier, 1994; Nosofsky & Palmeri, 1996) that may be thought to reflect the functioning of the implicit system (Ashby et al., 1998). Thus, it seems that depending on the researchers' theoretical commitments, the same learning paradigm has been argued to engage either explicit or implicit processes (see also Rehder & Hoffman, 2005, for suggesting that both explicit and implicit processes are recruited during the task). A modified version of the Type II task was created for the experiments presented here, by utilizing novel and arguably hard-to-name—rather than highly familiar—shapes. There was no evidence suggesting that names for the categorized items facilitate explicit processes of rule-discovery in Chapters 3 and 4. This result may be taken to suggest that the modified version of the Type II task does not recruit explicit rule-discovery learning processes. That is, names for the stimuli may indeed facilitate explicit category learning processes, but this effect was not observed in the solving of the modified version of the Type II category structure because learning was mediated by the implicit system. The experiments presented here were not designed to specifically examine the nature of learning processes (explicit vs. implicit) recruited during the modified version of the Type II task, and only further research with planned comparisons might elucidate this issue.

To conclude, it may be argued that changing the surface structure of a learning paradigm (in particular replacing familiar/nameable with novel/hard-to-name shapes) might severely affect learning processes in terms of representations or learning processes involvement, and care should be taken when generalizing assumptions across different versions of a learning paradigm.

Multiple plausible explanations were offered for the inability to replicate rule-discovery facilitation due to names for the categorization items (or their features) across learning paradigms and learning modalities. My experience though with investigating category learning, and in particular with using hard-to-name stimuli, is that the choice of perceptual modality is of great importance. Pilot experiments and participants' informal reports showed that participants faced less difficulty with naming visual stimuli (e.g., novel shapes or hard-to-name colors) compared to artificial tones. There seems to be a natural tendency to name visual—but not auditory—stimuli, thus naming manipulations might be less effective in the visual compared to the auditory modality. It could therefore be argued that re-examining the effect of names for the stimuli on auditory category learning might be proven a worthy endeavor. Maddox, Chandrasekaran, Smayda, & Yi (2013) recently provided a deterministic rule-based category structure in the auditory modality employing stimuli composed of four binary-valued dimensions (an analogue of the Type II task, see Figure 1A). I suggest applying the training manipulation presented here on the Maddox et al. structure and re-examining the effect of names for the features of auditory categorization items. In particular, I suggest using artificial hard-to-name tones instead of vowels and using stimulus identity as a category-diagnostic dimension (see Figure 1). Different groups of participants will be trained to associate the stimuli's sound identity with either verbal labels or hard-to-name ideograms. It may be predicted that names for the stimuli's features in an auditory category learning task might facilitate explicit learning processes.

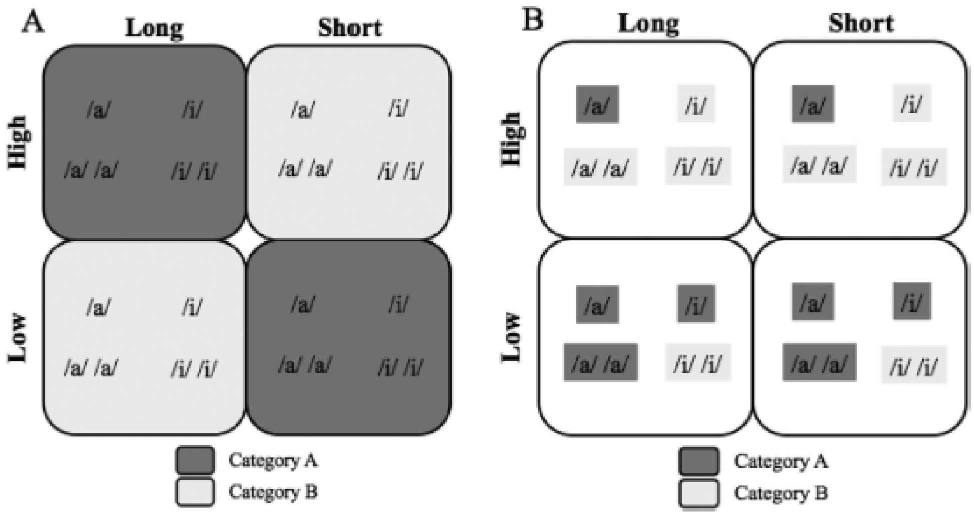


Figure 1. Schematic of a rule-based (A) and of an information-integration (B) category structure in the auditory modality. The categorization items are composed of four binary-values dimensions, namely pitch (low vs. high), duration (long vs. short), stimulus numerocity (one vs. two) and stimulus identity (/a/ vs. /i/). Here, pitch and duration are the category-diagnostic dimensions. From Maddox, Chandrasekaran, Smayda, & Yi (2013).

One of the open questions in the study of human category learning concerns the nature of the interaction between distinct learning systems. The COVIS model of category learning suggests that the two purported systems—the declarative/explicit and the procedural/implicit—might operate in parallel and that the most successful system—in terms of categorization accuracy—might assume response delivery (Ashby et al., 1998). Neuroimaging studies have provided novel insights on the matter, and suggested that the two systems might compete during learning. In particular, it has been suggested that the engagement of a learning system might inhibit the engagement of the other one (Moody, Bookheimer, Vanek, Knowlton, 2004; Poldrack et al., 2001). An alternative hypothesis is that the two systems might compete not during the acquisition, but rather during the application of knowledge (Foerde, Knowlton, & Poldrack, 2006). Recent behavioral investigations of the interaction of the two systems have provided mixed results in favor both of the independence and the competition between the systems (Ashby & Crossley, 2010). Thus further research is required to shed light on the interaction of the declarative and the procedural system of category learning.

In the literature of dual-systems theories of categorization, a fruitful approach in investigating the cognitive mechanisms of learning has been the experimental manipulation of the engagement of the purported systems. For example, researchers have manipulated the timing of feedback and stimulus-offset (Worthy, Markman, & Maddox, 2013), feedback informativeness (Maddox, Love, Glass, & Filoteo, 2008), secondary task demands (Miles & Minda, 2011; Zeithamova & Maddox, 2006), or the nature of the initial training regime (Ashby & Crossley, 2010). It was previously argued that re-examining the effect of names for the stimuli in auditory category learning might be proven a worthy endeavor. Insofar as names for the categorization items boost explicit processes of learning to categorize in the auditory modality, it may be suggested that the training manipulation presented here might be proven helpful in elucidating the interaction between the declarative and procedural system during an information-integration auditory categorization task (see Figure 1B for an example of an information-integration category structure).

Recent research has suggested that adult healthy participants approach an information-integration task through sub-optimal verbal rules which are later on replaced by implicit learning strategies (Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010). Moreover, Ashby and Maddox (2011) suggested that—during the learning of an information-integration task—the prevalence of the rule-based explicit strategies might prevent the transition to the procedural system (see also Ashby & Crossley, 2010, for evidence supporting this assumption). It may therefore be predicted that names for the categorization items' features may delay the transition from rule-based to information-integration strategies (as compared to associating features to hard-to-name visual symbols). Monitoring brain activity through neuroimaging techniques could furthermore elucidate the question of whether this delay is due to reduced engagement of the implicit system during initial learning, or rather due to the reduced access of the normal-functioning implicit system to the motor output system (Ashby & Maddox, 2011). Thus, the training manipulation of associating stimulus features to either verbal labels or hard-to-name visual symbols might be proven a useful tool in modulating the engagement of complementary learning systems, and thus in studying the nature of their interaction.

After examining the effect of verbal labels for categorization items (Chapter 2 and Chapter 3), the effect of verbal labels for the formed categories was investigated (Chapter 4). Various open questions need to be addressed concerning the initial and sustained effect of labels for the categories.

The label-feedback hypothesis (Lupyan, 2012a; 2012b) posits a mechanism of selective activation of diagnostic features that results in a label advantage during initial learning (Lupyan et al., 2007). Indeed, the label-advantage during category learning was replicated in a series of three experiments. Moreover, to examine if this label advantage is categorization-specific, learning to categorize was contrasted with learning to associate. Pooling the data from two experiments showed that the label advantage during learning to categorize was greater than during learning to associate. This result is—to my knowledge—the first piece of evidence suggesting that the mechanism of selective activation of diagnostic features might complement a domain-general verbal facilitation phenomenon (Tolins & Colunga, 2015). One possibility is that the facilitation during learning to categorize is the sum of two effects: the facilitation due to the selective-activation mechanism and the domain-general verbal facilitation. Another possibility is that the facilitation during learning to categorize exclusively stems from the categorization-specific mechanism of selective activation of diagnostic features. The experiments presented here are compatible with both accounts, so future research with planned comparisons should address this open question.

I also examined the sustained effects of labels for the formed categories on the processing of category-diagnostic perceptual features. In two experiments it was revealed that shapes that had previously been diagnostic of named categories captured attention to a greater extent compared to shapes that had previously been predictive of hard-to-name categories. This effect was found to be categorization-specific, in that following learning to associate there was no evidence of any difference in the processing of shapes that had previously been paired with labels or hard-to-name symbols. These results are novel (cf. Tolins & Colunga, 2015) and are equally compatible with perceptual learning accounts of category learning (Folstein, Palmeri, Van Gulick, & Gauthier, 2015; Goldstone, 1994), with learned attention accounts (Goldstone & Steyvers, 2001; Krushke, 1996), or with the label-feedback hypothesis (Lupyan, 2012a; 2012b). Further research (see Chapter 5 for suggestions) should elucidate the origin of the sustained effect of category labels on attention.

Finally, an important open question with respect to the initial and sustained effects of category labels is whether these effects are dimension-specific. Previous studies have suggested that labels for the categories lead participants to rely on dimensions that are typically category-diagnostic in real world situations, such as shape or frequency (Brojde, Porter, & Colunga, 2011; Perry & Lupyan, 2014). It could thus be argued that evolutionary processes have equipped some perceptual dimensions with a special status or privilege regarding their role in learning to categorize. The mechanism of selective activation of diagnostic features (Lupyan, 2012a; 2012b) does not seem to account for this purported special status of some perceptual dimensions. Future research should examine the initial and sustained effects of category labels utilizing typically non-diagnostic dimensions. If the effects of category labels are found to be dimension-specific, then the label-feedback hypothesis should be modified to account for this purported dimension specificity.

Conclusions

An important question regarding the architecture of cognition concerns the extent to which cognitive faculties and their corresponding representations are encapsulated (e.g, Fodor, 1985) or rather interact with other cognitive faculties. Recent theoretical and empirical accounts have brought about a renewed interest on the Whorfian hypothesis (Gleitman & Papafragou, 2013; Regier, Kay, Gilbert, & Ivry, 2010) that language modulates faculties such as color perception or category learning (e.g., Regier & Kay, 2009; Lupyan et al., 2007, but see Firestone & Scholl, 2015, for an opposing account).

The present dissertation contributes to the language and thought debate by suggesting that verbal labels for the categorization items and also for the formed categories affect category learning and attention processes. Notwithstanding the fact that some of the described findings call for replication or further investigation, it may be argued that the experiments presented here contribute to theories of learning by suggesting a special role for verbal labels in learning to categorize.

References

Ashby, F. G, Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442-81. doi:10.1037/0033-295X.105.3.442

Ashby, F.G., & Crossley, M.J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory, 94*(1), 1–12. doi:10.1016/j.nlm.2010.03.001

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*(1), 147–161. doi:10.1111/j.1749-6632.2010.05874.x

Brojde, C. L., Porter, C., & Colunga, E. (2011). Words can slow down category learning. *Psychonomic Bulletin & Review, 18*(4), 798–804, http://dx.doi.org/10.3758/s13423-011-0103-z

Chandrasekaran, B., Yi, H. G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review, 21*(2), 488–495. doi:10.3758/s13423-013-0501-5

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology, 21*, 561–580. doi:10.1080/09541440802097951

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 20*, 1–77. doi:10.1017/S0140525X15000965

Fodor, J. A. (1985). Précis of the modularity of mind. *Behavioral and Brain Sciences, 8*(1), 1–42. http://dx.doi.org/10.1017/S0140525X0001921X

Foerde, K., Poldrack, R. A., Knowlton, B. J. (2007). Secondary task effects on classification learning. *Memory & Cognition, 35,* 864–874. doi:10.3758/BF03193461

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Science, 103,* 11778–11783. doi:10.1073/pnas.0602659103

Folstein, J. R., Palmeri, T. J., Van Gulick, A. E., & Gauthier, I. (2015). Category learning stretches neural representations in visual cortex. *Current Directions in Psychological Science, 24*(1), 17–23. http://dx.doi.org/10.1177/0963721414550707

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178–200. doi:10.1037/0096-3445.123.2.178

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*(1), 116–139. http://dx.doi.org/10.1037/0096-3445.130.1.116

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106–120. doi:10.1101/lm.1.2.106

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22–44. http://dx.doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science, 8*(2), 225–248. http://dx.doi.org/10.1080/095400996116893

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition, 3*, 54. doi:10.3389/fpsyg.2012.00054

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Waltham, MA: Academic Press, doi:10.1016/B978-0-12-394293-7.00007-8

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18*, 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H. G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging, 28*(4), 1042–1056. doi:10.1037/a0034969

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition, 108*(2), 578–589. doi:10.1016/j.cognition.2008.03.010

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia, 48*(10), 2998–3008. doi:10.1016/j.neuropsychologia.2010.06.008

Miles, S. J., & Minda, J. P. (2011). The effect of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37*, 588–607. doi:10.1037/a0022309

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and nonrule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1518–1533. doi:10.1037/a0013355

Minda, J. P., & Miles, S. J. (2010). *The influence of verbal and nonverbal processing on category learning*. In B. H. Ross (Ed.), The Psychology of learning and motivation: Advances in research and theory, 52, 117–162. San Diego: Academic Press. doi:10.1016/S0079-7421(10)52003-6

Moody, T. D., Bookheimer, S. Y., Vanek, Z., Knowlton, B. J., 2004. An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. *Behavioral Neuroscience, 118*, 438–442.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22*(3), 352–369. http://dx.doi.org/10.3758/BF03200862

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral dimension stimuli. *Psychonomic Bulletin & Review, 3*(2), 222–226. http://dx.doi.org/10.3758/BF03212422

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language, 135*, 66–72. http://dx.doi.org/10.1016/j.bandl.2014.05.005

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso, J., Myers, C. E., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546-550. doi:10.1038/35107080

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory system debate. *Neuroscience and Behavioural Reviews, 32,* 197–205. doi:10.1016/j.neubiorev.2007.07.007

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 13*(10), 439–446. doi:10.1016/j.tics.2009.07.001

Regier, T, Kay, P., Gilbert, A. L., & Ivry, R. B. (2010). Language and thought: Which side are you on anyway? In B. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 165–182) . Oxford, England : Oxford University Press.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*(1), 1–41. http://dx.doi.org/10.1016/j.cogpsych.2004.11.001

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196. doi: 10.1006/jmla.2001.2839

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42. doi:10.1037/h0093825

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in Rhesus Monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General, 133*(3), 398–414.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science, 19*(11), 1169–1177.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8*, 168–176. doi:10.3758/BF03196154

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2013). Feedback and stimulus-offset timing effects in perceptual category learning. *Brain & Cognition, 81*, 283–293. doi:10.1016/j.bandc.2012.11.006

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34*(2), 387–398. doi: DOI: 10.3758/BF03193416

Appendix

Summary

Recent theories of cognition argue against a modular cognitive architecture and instead suggest that language interacts with presumably non-linguistic faculties such as learning or perception. In the present dissertation I investigated the interplay between language and category learning, with a specific focus on verbal labels. In particular, I examined the effect of verbal labels for (a) the to-be-categorized material, and (b) the formed categories on learning to categorize.

Based on dual-systems theories of category learning, it was reasoned that participants would find it more easy to develop and apply explicit rules of category membership when the to-be-categorized material is readily nameable compared to hard-to-name material. To test the idea of rule-discovery facilitation due to names for the stimuli, separate groups of participants were trained for three consecutive days to associate hard-to-name auditory tones to verbal labels or hard-to-name ideograms; or to associate cue identity to hard-to-name colors; a forth group remained unexposed to the cues (Chapter 2). Following training, participants were administered an auditory version of a probabilistic category learning task, utilizing the trained tones as cues. Results revealed that the verbal-label group outperformed the ideogram group, suggesting that names for the auditory cues facilitated rule-discovery learning processes. It was additionally revealed that the ideogram group outperformed the intensity-training group and that the intensity-training group outperformed the no-training group, suggesting that cue individuation and familiarity with the stimuli facilitated explicit processes of learning to categorize.

The idea of rule-discovery facilitation due to names for the stimuli was further tested by employing a deterministic category learning task in the visual modality (Chapter 3). Separate groups of participants were trained for two consecutive days to associate hard-to-name shapes to verbal labels or hard-to-name ideograms; a third group received mock training and remained unexposed to the shapes. Following training, participants were administered a category learning task with categorization items composed of the trained shapes. Results revealed no group-level differences, there was thus no evidence suggesting that names for the features of categorization items facilitate explicit categorization processes. The inability to replicate the effect of rule-discovery facilitation due to names was interpreted as suggesting that changing the surface structure of a learning paradigm

severely interferes with the task's processing demands in terms of representations and learning system engagement.

The next research endeavor of the present dissertation examined the effect of labels for the formed categories (Chapter 4). The label-feedback hypothesis posits a mechanism of selective activation of diagnostic features due to labels for the categories. This mechanism is predicted to provide a label advantage during learning and also affect attention in the long term. To test the label advantage during initial learning, participants were trained to learn named and hard-to-name categories. The sustained effect of category labels were examined by employing the category-diagnostic perceptual features in posttraining test tasks, and in particular in category learning tasks, and in a visual discrimination task using eye tracking. To examine if the selective-activation mechanism is categorization specific, learning to categorize was contrasted with learning to associate, by training control groups of participants to learn named and hard-to-name association instead of categories. Results suggested that named categories were consistently learned more accurately than hard-to-name categories. This label advantage during learning to categorize was found to be greater compared to a label advantage during learning to associate, supporting the notion of a categorization-specific mechanism offering facilitation. With respect to the sustained effects of labels, category-diagnostic perceptual features of named categories were found to capture attention to a greater extent compared to diagnostic features of hard-to-name categories. This long-term effect of labels on attention was categorization-specific, in that following learning to associate there was no evidence suggesting differential processing of diagnostic features of named compared to hard-to-name associations. These results were collectively interpreted as supporting the label-feedback hypothesis of a categorization-specific mechanism of selective activation of diagnostic features due to names for the formed categories.

The present dissertation examined the effects of labels for the categorization items and for the formed categories on processes recruited during and also following learning. The results presented here suggest that verbal labels affect category learning and attention processes, and thus support the notion that language in not merely an interface serving communication purposes but rather modulates—or perhaps shapes—cognition.

This page was intentionally left blank.

Acknowledgments

I would like to express my gratitude to Prof. Athanassios Protopapas for accepting to be my supervisor. His integrity, his scientific thinking, his focus on work, and his devotion to the painstaking endeavor of exploring the principles and mechanisms of cognition, were all sources of inspiration to me. Thanassi has been a great mentor, and a concerned and caring guide. I would also like to thank Argiro Vatakis for being constantly supportive, and for always believing in me. Petros Roussos was a great teacher and offered invaluable help by urging his students take part in our lab's experiments; I couldn't thank him enough. This journey was accompanied by an unanticipated gift: the friendship of Laoura Ziaka, whom a admire greatly as a scientist and value even more as a friend. Laoura was a true partner: we set up and tested the eyetracking equipment together, we organized student participation for our experiments, we discussed and formulated hypotheses, we spent some time together in a hospital room last November. This dissertation is Laoura's too. Sonia Loui was also a beloved and respected colleague who I greatly thank her for her support and friendship. Angeliki Andrikopoulou, Lora Dimopoulou, Efi Kapnoula, Litsa Xeimariou, it's been great being a member of the Language and Learning Lab alongside you. Petros, thank you for your support and for respecting my devotion to my studies, thank you for everything.

This dissertation is dedicated to my parents, who were willing to sacrifice a lot in order to support me both morally and financially throughout the years of my doctoral studies. I owe them both my being and my well-being.