No Effect of Verbal Labels for the Shapes on Type II Categorization Tasks

Fotis A. Fotiadis (f_fotis@phs.uoa.gr)
Department of Philosophy & History of Science, University of Athens, Ano Ilissia Campus
GR-157 71, Athens, GREECE

Athanassios Protopapas (aprotopapas@phs.uoa.gr)
Department of Philosophy & History of Science, University of Athens, Ano Ilissia Campus
GR-157 71, Athens, GREECE

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.

Keywords: Verbal labels; hypothesis testing; categorization; learning

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[1], in an exclusive-or fashion. A simple verbal rule seems to be able to define category membership (e.g., “black triangles and 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

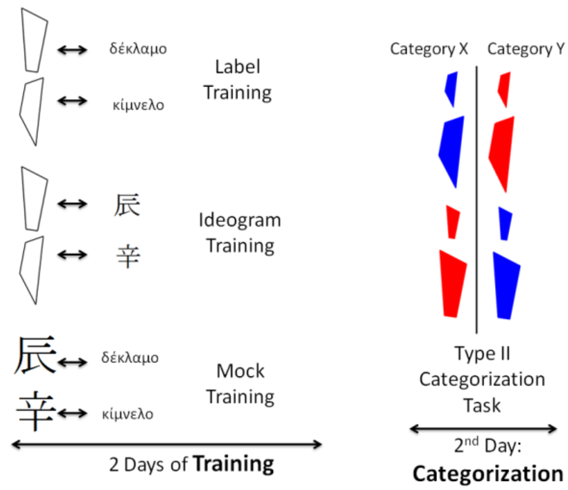

Figure 1: Design of the present study.

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).

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

[1] 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.

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

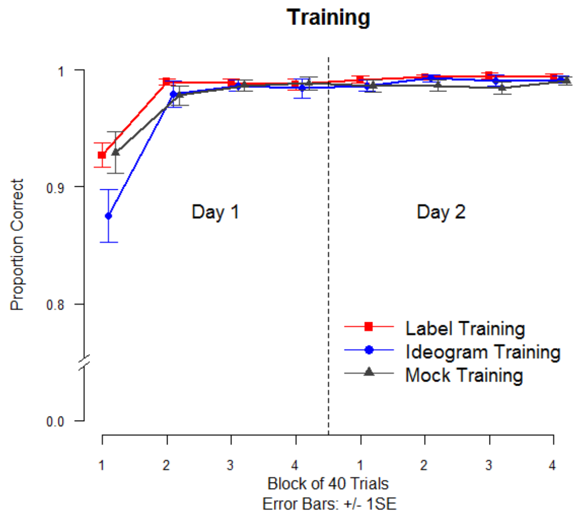

Figure 2: Learning curves in the three training conditions in blocks of 40 trials

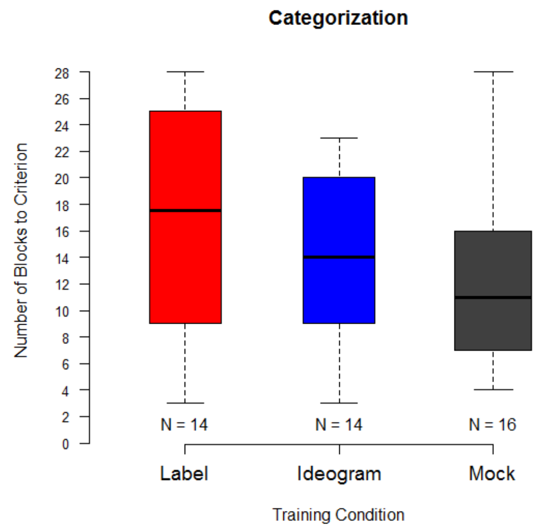

Figure 3: Number of blocks to reach learning criterion in the categorization task, per training condition. Data from learner participants only. Boxes denote interquartile range; thick lines mark the median; error bars extend to the full range; N denotes sample size.

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% (*SD*= 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.

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 (χ^2 = .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[2] (see Fig. 3). An one-

[2] 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

way analysis of variance revealed that there was no effect of training condition on the number of blocks to reach the learning criterion, $F(2, 42) = 1.777$, $\eta^2 = .080$, $p = .182$.

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

---

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.

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.

Acknowledgments

We thank Laoura Ziaka for help with collecting data.

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.