Drivers of Incidental Category Learning

Lori L. Holt (loriholt@cmu.edu)[1,2], **Casey L. Roark** (croark@andrew.cmu.edu)[1,2], **Matt I. Lehet** (lehetmat@msu.edu)[1,2], and **Frederic Dick** (f.dick@bbk.ac.uk)[3]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15232 USA
[2]Center for the Neural Basis of Cognition, Pittsburgh, PA 15232 USA
[3]Department of Experimental Psychology, University College London, London, WC1H 0AP UK

Abstract

Little is known about how categories are learned *incidentally* without instructions to group objects, overt decisions about category identity, or feedback about these decisions. Here we investigate how category learning may occur based on the association of categories with behaviorally-relevant events and actions. Previous research developed the Systematic Multimodal Associations Reaction Time (SMART) task in which participants report the location of a visual target with a keypress. The location of an upcoming visual target is predicted by the identity of a novel sound category, exemplars of which precede appearance of the visual target. This category-to-location mapping supports incidental learning of auditory categories, with generalization to novel exemplars. Here, we examined whether this learning is driven by the category-to-location relationship, or instead by the association with distinct response alternatives. Across two experiments, we observe that both a covert, reaction time measure of category learning and an overt labeling task testing generalization of learning converge to indicate that the category-to-response relationship drives incidental learning in the SMART task.

Keywords: incidental learning; category learning; auditory

Introduction

Everyday behaviors - like identifying a series of beeps as a fire alarm or a cell phone ring, recognizing fruits as edible or spoiled, and deciding whether an unleashed dog is friendly or fierce - rely on categorization. The ability to treat distinct perceptual experiences as functionally equivalent is a vital component of human cognition. Although there is a rich literature on category learning, our understanding is largely based on laboratory studies conducted with visual objects and across training paradigms that involve explicit categorization. Typically, participants are aware that the objects should be sorted, they make overt category decisions, and receive feedback that directs their future decisions. This classic approach has provided a rich and informative literature characterizing category learning (Ashby & Maddox, 2005 for review). Nonetheless, results obtained across overt training with visual objects may not generalize, broadly, across other perceptual modalities or in natural environments that do not provide explicit training.

Indeed, category learning in natural environments often occurs under less explicit conditions, without instructions to search for category-diagnostic dimensions, overt category decisions or experimenter-provided feedback. These differences from category learning in the lab create a disconnect with ecological validity and may be important in that the stimulus structure, task, and timing of feedback are known to have an important influence on the mechanisms that are recruited for category learning (e.g., Ashby, Maddox, & Bohil, 2002; Maddox & David, 2005).

In the auditory domain, there has been some progress in developing new approaches to studying category learning with tasks that do not involve overt category decisions or explicit feedback about these decisions. In these *incidental category learning* studies, sound categories are learned by virtue of their relationship to success in performing a task defined along other, largely visuomotor, dimensions. Although participants do not overtly search for dimensions diagnostic to category membership and do not receive explicit feedback, this learning is neither passive, nor entirely unsupervised or feedback-free (Gabay et al., 2015; Lim et al. 2011; see Seitz et al., 2010; Vlahou et al. 2012). Incidental learning is thus distinct from unsupervised learning in that feedback is present. Unlike explicit category learning, this feedback is not directly related to category decisions, but rather is provided via elements that are statistically associated with the categories.

For example, in a task developed by Wade and Holt (2005), the objective is to navigate a space-themed videogame environment, targeting approaching aliens with a laser. Participants are instructed only in how to maneuver in the game. They are not overtly encouraged to form audio-visual or audio-motor associations and they are not told the significance of the sounds, which are embedded in a more complex soundscape. The videogame task is largely visuomotor, but it is organized in such a way that sound category learning can support successful navigation. Specifically, each alien creature is associated with multiple, acoustically-variable sounds drawn from an auditory category. When an alien appears in the videogame, an associated sound-category exemplar is repeatedly played. As players advance to higher levels, the pace of play becomes more challenging and there is increasing opportunity for the sound categories to support behavior in the primary game navigation task because participants can *hear* an approaching alien before *seeing* it. Thus, there is an advantage in learning to categorize across the acoustically-variable sounds associated with specific aliens. Indeed, participants quickly learn both novel artificial nonspeech auditory categories (Wade & Holt, 2005; Lim, Lacerda, & Holt, 2015) and also non-native speech categories (Lim & Holt, 2011; Lim et al. 2015) and generalize this learning to novel category exemplars in a post-game overt labeling task in which novel sounds are matched with alien creatures. Successful incidental auditory category learning engages putatively speech-selective left posterior superior temporal cortex for newly-learned nonspeech categories (Leech, Holt, Devlin & Dick, 2009) and warps perceptual space in a manner akin to speech category acquisition (Liu & Holt, 2011).

Although other-worldly, this task's demands more closely approximate those of learning in a natural environment than traditional explicit or passive-exposure learning paradigms because sound categories are associated with behaviorally-relevant events and actions. This laboratory-based paradigm captures some of the incidental nature of learning categories in more natural environments. Yet, a trade-off is that it is difficult to uncover which of its many elements are responsible for driving learning.

Recognizing this constraint, Gabay, Dick, Zevin, and Holt (2015) developed a highly simplified task that includes some aspects of the videogame that may be important in driving incidental learning. In the Systematic Multimodal Association Reaction Time (SMART, Figure 1) task participants rapidly detect the appearance of a visual target in one of four possible screen locations and report its location by pressing a key corresponding to the visual screen position.

Critically, a brief sequence of sounds precedes each visual target. Unknown to participants, the sounds are drawn from one of four distinct sound categories (Figures 1a, 2). There is a multimodal (auditory-category to visual-location) correspondence that relates variable sound category exemplars to a consistent location into which a visual object will appear (Figure 1b). This mapping is many-to-one, such that multiple, acoustically-variable sound category exemplars are associated with a single visual location. Likewise, sound categories are predictive of the *action* required to complete the task. In the training blocks (Figure 1d), the categories perfectly predict the location of the upcoming visual detection target and the corresponding response button to be pressed. Thus, learning to treat the acoustically-variable sounds as functionally equivalent in predicting the upcoming visual target location may facilitate visual detection without requiring overt sound categorization decisions or even awareness of category structure. Participants are not instructed about the utility of the sounds and the many-to-one association of sounds to locations prevents simple auditory-visual associations from driving behavior.

Category learning can be measured covertly online during the SMART task because the fourth block of trials destroys the association between auditory category and visual location (Figure 1d). If participants incidentally learn sound categories to support quick detection of the visual target then we expect visual detection to be *slower in the test block relative to the training block that preceded it* – a reaction time cost (RT Cost). Additionally, an overt sound categorization post-test follows the SMART task. In the post-test, participants hear novel sound exemplars drawn from the sound categories and guess the location where the visual target would be most likely to appear; no visual targets appear and there is no feedback about the correctness of responses. This provides a measure of generalization, a hallmark of

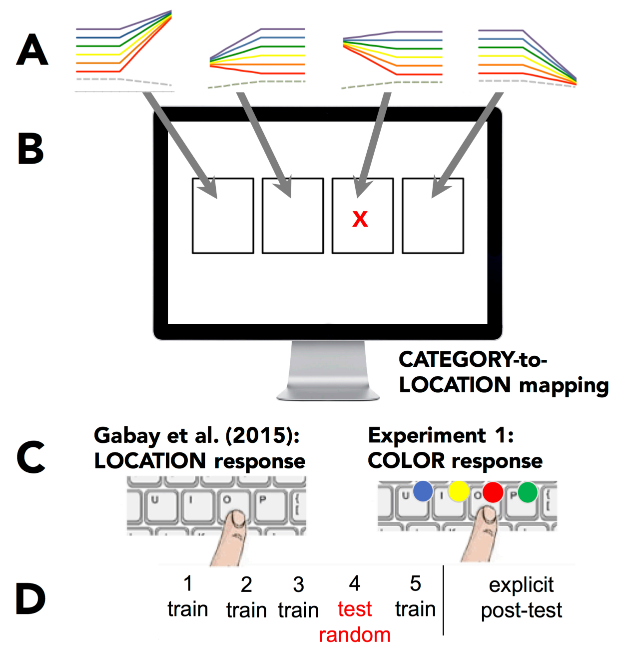

Figure 1. Overview of SMART Paradigm. (A) Four auditory categories are defined by multiple exemplars, see Figure 2. (B) Each category is associated with a particular visual target location. (C) In Gabay et al., (2015), participants indicate the target location with a key press. In Experiment 1, participants indicate target color with a keypress. (D) Blocks include a Test Block in which the category-to-location association is destroyed, and an overt labeling post-test follows SMART training.

robust category learning. It also requires that participants transfer learning to an explicit task that differs from the incidental SMART learning context.

Note that although SMART shares some characteristics with a traditional procedural learning paradigm, the serial reaction time (SRT) task (Nissen & Bullemer, 1987), SMART – at least as described above – measures category learning, not sequence learning; there are no embedded sequences. SRT and SMART share the fact that participants are not alerted to a regularity in the training stimuli. This is done to promote incidental learning conditions, although there is a good deal of debate about the extent to which such tasks may be implicit (Shanks, 2003).

Gabay et al. (2015) examined nonspeech auditory category learning in the simplified SMART paradigm across the same sound exemplars employed by Wade and Holt (2005) in the incidental videogame paradigm. Although the task was a simple visual detection, participants nonetheless learned the auditory categories. Destroying the association between the sound categories and the upcoming location of the visual target resulted in a significant reaction time cost, indicative of a reliance on auditory categorization to facilitate speedy visual target detection. Moreover, this learning generalized to labeling novel auditory exemplars in the post-training test. Here, we pursue the implications of the Gabay et al. (2015)

results to better understand the factors that drive incidental category learning. Specifically, we explore the necessity of the category-to-location correspondence (Experiment 1) and the association of the categories with distinct response alternatives (Experiment 2).

Experiment 1

In the SMART task category membership of the sounds presented prior to the target predict the location of the upcoming visual target (Gabay et al., 2015). Participants report the specific location of the visual target with a unique keypress such that each of four fingers on the dominant hand is associated with a unique visual location. This establishes a situation in which the auditory categories are associated with the visual characteristics of a trial (target location), as well as the response (finger reporting location). Experiment 1 tests which of these factors supports incidental category learning by decoupling the link between the category-location association of auditory-visual stimuli and the category-response association of auditory categories to motor responses.

Methods

Participants Twenty-four young adult participants were recruited from Carnegie Mellon University or the Pittsburgh community. They received course credit or a small payment for their time. All participants had normal or corrected-to-normal vision and reported normal hearing.

Stimuli The auditory categories were defined by nonspeech sound exemplars identical to those used by Gabay et al. (2015), as originally developed by Wade and Holt (2005) and illustrated in Figure 2. These sounds have some of the spectrotemporal complexity of speech but are unequivocally nonspeech owing to their noise and square wave sources. Six unique exemplars from each category were used in SMART training; an additional 5 novel exemplars were withheld from training for use in testing generalization of category learning at post-test. Two categories were defined by a simple acoustic cue (up- or down-sweep in frequency of a higher-frequency component, Figure 2a and 2b). The other two categories were defined in a more complex, higher-dimensional perceptual space (no single acoustic cue uniquely defined category membership, Figure 2c and 2d). Each exemplar was 250 ms in duration and exemplars were matched in RMS amplitude.

Procedure In large part, the procedure followed the SMART paradigm of Gabay et al. (2015), as described above. As in Gabay et al. the sound categories predicted the upcoming *location* of the visual target (as in Figure 1b). However, instead of a single visual target (red 'X') as in Gabay et al., Experiment 1 included four distinct visual targets distinguished by color (red, blue, green, and yellow 'X' targets). Instead of responding to indicate target *location*, participants responded to the target *color* using a standard keyboard (*u*, *i*, *o*, *p* keys) with colored stickers to facilitate the color-response mapping (Figure 1c). Each of the

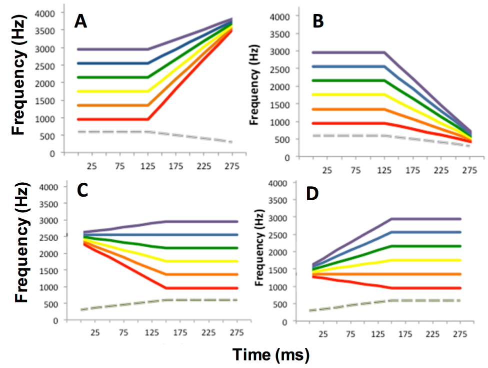

Figure 2. Auditory Categories. Each higher-frequency (colored) component is paired with the lower-frequency (grey) component to create 6 category exemplars for training. The 5 generalization exemplars are not pictured.

distinctly-colored targets appeared in each location with equal probability. Whereas sound categories perfectly predicted the *location* of an upcoming target, they did not predict the appropriate *color response*.

This manipulation decoupled the association of auditory-category to visual-location from the response mapping. If an auditory-visual association (category-to-location) is sufficient to drive incidental category learning then we expect to observe learning compatible with the results of Gabay et al. However, if the mapping to the overt response is an important contributor to learning, then the decoupling of auditory categories to response should eliminate or reduce incidental auditory category learning.

On each trial, five distinct 250-ms sound exemplars drawn from one of the four auditory categories were presented (0 ms ISI, 1250 ms total duration) preceding presentation of the visual target. Thus, training trials involved within-category variability of sound exemplars, similar to an approach producing robust learning in the Gabay et al. (2015) studies. Reaction time was measured from the onset of the visual target to the keypress.

After 8 practice trials (in which sound categories did not predict target location), participants completed 5 blocks of the SMART task. The first three blocks were each composed of 96 trials (4 categories x 6 exemplars x 4 repetitions) for which the preceding sound exemplars' category membership perfectly predicted the location at which the visual target would appear (but not the color response). The location-to-category relationship was destroyed in the fourth block; sound categories were randomly assigned to visual target location across 48 trials. In the final, fifth, 96-trial block the category-to-location association present in the first three blocks was restored. Participants are not alerted to the relationship of the sounds to the task and the acoustic variability among within-category sound exemplars assured that there was no simple sound-location or sound-color association. Participants were encouraged to take brief, self-paced breaks between blocks.

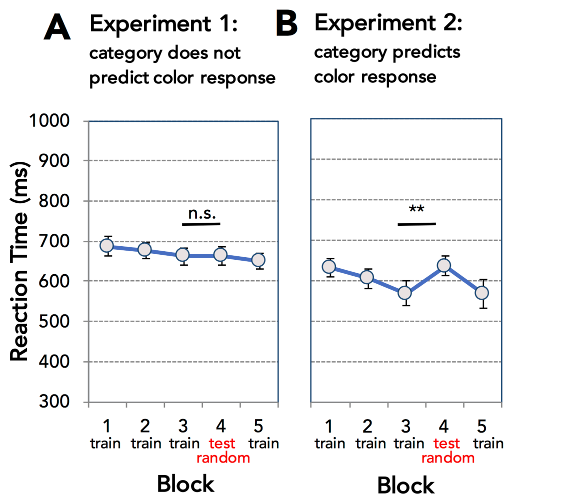

Figure 3. Target Detection RT. (A) Experiment 1, no RT Cost. (B) Experiment 2, RT Cost. Error bars are standard error of the mean.

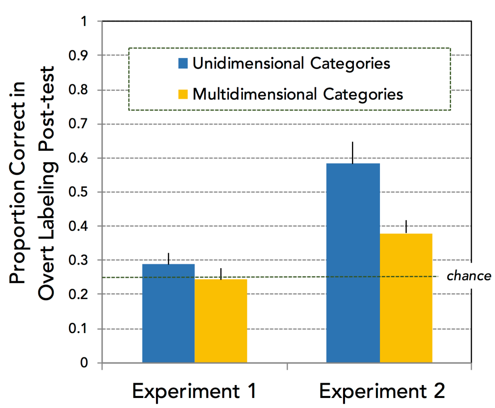

Figure 4. Mean accuracy in the overt labeling post-test. Plotted as a function of unidimensional versus multidimensional categories in Experiment 1 and Experiment 2. Chance performance is 0.25, error bars are standard error of the mean.

Results

Reaction Time (RT) Cost Examination of reaction time to respond to the visual target's color in Block 4, within which the relationship of the sound category to the visual location experienced across the first three blocks is disrupted by randomization, compared to Block 3 provides a covert measure of the extent to which categorization supports performance in the SMART task, without requiring an overt categorization decision. If participants learned the categories sufficiently to predict the location of the upcoming target, they should be slower during the random Block 4.

Trials for which there was a visual detection error (4.6% of trials) or reaction times shorter than 100 ms or longer than 1500 ms (2.7% of trials) were excluded from analyses.

Figure 3a plots RT as a function of SMART task block. Consistent with what is apparent visually, there was no significant RT Cost from Block 3 to Block 4 ($t(23)=0.126$, $p=.901$; mean Block 3, 662.1 ms, S.E.=21 ms; mean Block 4, 662.8 ms, S.E.=23 ms). Category learning was not evident in the RT Cost measure.

Overt Labeling In a similar manner, there was no evidence of category learning in the overt labeling task. Participants were unable to guess the location of an upcoming target based on sound category exemplars. As shown in Figure 4, labeling accuracy was no different from chance, $t(23)=.531$, $p=.60$.

This was true for both categories defined by a simple unidimensional cue, $t(23)=1.094$, $p=.285$, and also for more the categories defined by a more complex, multidimensional pattern of cues, $t(23)=-.192$, $p=.849$.

Summary

Gabay et al. (2015) reported robust incidental auditory category learning across the same stimuli and largely the same SMART paradigm employed here. Participants in Experiment 1 experienced the same category-to-location relationship effective in driving incidental category learning in the Gabay et al. studies. Only the nature of response, differed. In Experiment 1, decoupling the response (to color) from the predictive category-to-location association resulted in no evidence of category learning the covert or overt tasks.

Experiment 2

The results of Experiment 1 suggest that the alignment of the response with the auditory-visual association is essential in driving incidental category learning. If this is the case, then incidental auditory category learning should be reinstated by aligning the auditory-visual association with the color response. To test this, participants in Experiment 2 responded to the color of the visual target, as in Experiment 1. However, here, the auditory categories predicted visual target *color* (and therefore response) rather than *location* (see Figure 5). We predict that reinstating the predictive association between auditory categories and response along the color dimensions will result in incidental auditory category learning.

Methods

Participants Twenty-one young adults, recruited in the manner of Experiment 1, participated.

Stimuli The stimuli were identical to those of Experiment 1.

Procedure The procedure was the same as Experiment 1, except that the sound categories predicted which *color* would be associated with the upcoming visual target instead of which *location* would be associated with the target. Participants responded based on color and so, unlike Experiment 1, the auditory-visual association was aligned with response, as in Gabay et al. (2015).

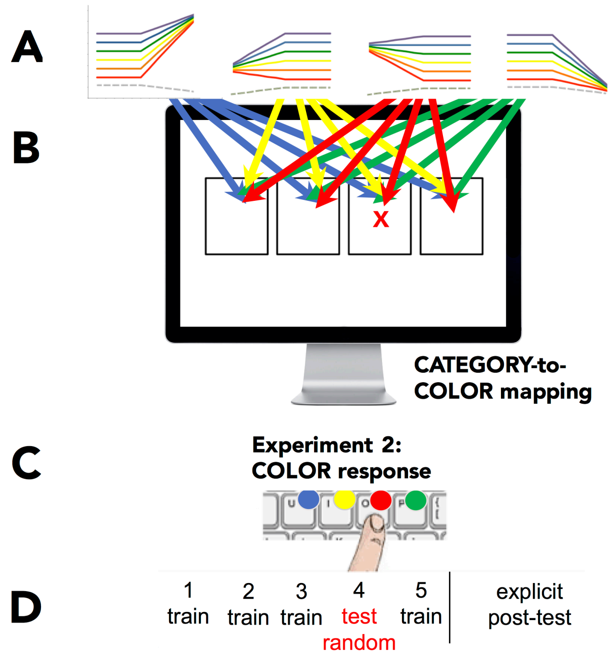

Figure 5. Paradigm, Experiment 2. (A) Four auditory categories as in Gabay et al. (2015). (B) Here, each category is associated with a particular visual target color (blue, yellow, red, green) appearing equally often at each location. (C) In Experiment 2, participants indicate target *color.* (D) Training and testing as Experiment 1.

Results

Reaction Time (RT) Cost Figure 3b plots RT as a function of SMART task block. Trials for which there was a visual detection error (4.5% of trials) or for which reaction time was shorter than 100 ms or longer than 1500 ms (2.4% of trials) were excluded from analyses. In contrast to Experiment 1, there was a significant RT Cost from Block 3 to Block 4, $t(20)=2.66$, $p=.015$; mean Block 3, 567.4 ms, S.E. 30 ms; mean Block 4, 625.5 ms, S. E., 23 ms).

Overt Labeling The right panel of Figure 4 plots the proportion of correct responses in the overt labeling task. In contrast to Experiment 1, participants successfully generalized category learning from the SMART task to label novel sound category exemplars according to the expected color of a visual target, $t(20)=4.420$, $p=.00026$. Accuracy was above chance for both unidimensional categories, $t(20)=4.73$, $p=.0001$, and multi-dimensional categories, $t(20)=3.294$, $p=.004$.

Summary Experiments 1 and 2 were highly similar. Participants experienced the same sound category exemplars, with the same number of trials and repetitions. They performed exactly the same visual task – reporting the color of the visual target – under the same task demands. Yet, whereas participants in Experiment 1 exhibited no auditory category learning or generalization, participants in Experiment 2 exhibited evidence of learning in both covert and overt measures of categorization. What differed across experiments was the relationship of participants' response (visual target color identification) to the auditory categories. In Experiment 1, auditory categories predicted visual target location, but not the response dimension color, and led to no learning. When auditory categories predicted the visual target dimension associated with response, learning was evident in Experiment 2. Thus, incidental auditory category learning appears to be driven more so by the relationship of auditory categories to responses than of experienced statistical regularities across auditory-visual modalities.

General Discussion

Category learning is a central cognitive process required in everyday behaviors. It involves learning to treat perceptually distinct objects and events as functionally equivalent. In contrast to overt training in laboratory studies, category learning in the world often proceeds under conditions in which learners do not have instructions to search for category-relevant information, do not make overt category decisions, and do not experience feedback directly. Complementing earlier studies (Wade & Holt, 2005; Lim & Holt, 2011; Liu & Holt, 2011; Gabay et al. 2015), the present results emphasize that participants can *incidentally* learn perceptual categories as they undertake seemingly unrelated tasks, if the task demands of the primary task align with the structure of the categories.

The differences in learning observed across Experiments 1 and 2 help to delineate the nature of these task demands. Participants in Experiments 1 and 2 had largely the same experience. The stimuli were identical and presented equally often thereby equating stimulus experience. The training protocol was nearly equivalent. The mode of response was identical, as was the visual display of targets. Yet, when the behavioral responses were decoupled from the category-to-location association experienced in the primary visual detection task, there was no learning. Reinstating this coupling by introducing category-to-color association and requiring color responses led to learning.

This begs the question of whether unique category-to-response mappings (e.g., distinct fingers on distinct keys) are fundamental to binding acoustically variable sounds together to form new categories through incidental learning. Such a view would resonate with dual systems theories of category learning across explicit training that posit engagement of an implicit, or reflexive, system that is sensitive to response mappings (Ashby & Maddox, 2005). Even so, it is important to point out that Wade and Holt (2005) observed robust incidental category learning across the same auditory categories as the present study in a videogame paradigm that did not involve distinct motor responses for each category. Thus, unique mappings may support incidental auditory category learning, without being obligatory. It is possible that additional associations of the categories to behaviorally-relevant actions in the videogame supported learning in that paradigm. Further research will be needed to characterize additional contributors.

Finally, it is useful to consider these results in the context of passive, 'statistical' learning. Learning via mere exposure to distributions of category exemplars has often been taken to be a more ecologically-realistic alternative to learning from overt training with feedback and, indeed, under some circumstances category learning appears to proceed via mere exposure (e.g., Maye, Werker, & Gerken, 2002). In this regard, the substantial differences in learning observed across Experiments 1 and 2 are notable since the studies had highly similar tasks and shared identical stimuli. Observance of learning in Experiment 2, but not in Experiment 1, argues against the possibility that the category learning observed in Experiment 2 arose from passive exposure to the exemplars. This conclusion aligns with findings from previous studies utilizing the same nonspeech sounds as the present experiments, but in passive-exposure, 'statistical learning' paradigms (Emberson, Liu, & Zevin, 2013; Wade & Holt, 2005). In these studies, the multidimensional categories well-learned in Experiment 2 were not acquired via passive exposure, at least for exposure durations on par with the duration of training examined in the present experiments.

In this regard, incidental learning may be lie between overt training with feedback and mere exposure. It is active, rather than passive, but does not involve overt category decisions or feedback about categorization responses. Just as, in real-world environments, learners are typically active and can capitalize on rich multimodal associations existing between category exemplars with other objects and events, and their own behaviors, the supportive associations involved in incidental learning tasks may serve to hasten category learning that is difficult through mere exposure.

The SMART paradigm (Gabay et al. 2015), and even the videogame it models (Wade & Holt, 2005), are highly simplistic compared to the supportive multimodal correlations potentially available in the natural perceptual world. But, the present results suggest that the presence of co-occurring visual referents and actions may support category learning in the context of auditory category learning by signaling the distinctiveness of acoustically-similar items across referents or the similarity of acoustically-distinct exemplars paired with the same referent. Gabay et al. (2015) referred to these associations as the 'representational glue' that binds exemplars together in category learning. The present studies further clarify this by demonstrating that it is the association of category exemplars with a consistent *response* that is effective in supporting category learning.

Acknowledgments

The authors acknowledge the contributions of Christi Gomez in collecting the data. This work was funded by the National Institutes of Health (R01DC004674).

References

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

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*, 666–677.

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, *128*, 82–102.

Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 1124-1138.

Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *The Journal of Neuroscience, 29,* 5234–5239.

Lim, S. -J., Lacerda, F., & Holt, L. L. (2015). Discovering functional units in continuous speech. *Journal of Experimental Psychology: Human Perception & Performance, 41*, 1139-1152.

Lim, S.-J., Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*, 1390–1405.

Liu, R. & Holt, L. L. (2011). Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. *Journal of Cognitive Neuroscience, 23*, 683–698.

Maddox, W. T. & David, A. (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.

Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1–32.

Seitz, A. R., Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. *Vision Research*, *49*, 2604–2610.

Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General, 141*, 363-381.

Wade, T. & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America, 118*, 2618-2633.