The representational glue for incidental category learning is alignment with task-relevant behavior

Casey L. Roark[a,b], Matthew I. Lehet[a,b], Frederic Dick[c,d], & Lori L. Holt[a,b,e]

a. Carnegie Mellon University, Department of Psychology

b. Center for the Neural Basis of Cognition

c. Birkbeck College, University of London, Department of Psychological Sciences

d. University College London, Experimental Psychology

e. Carnegie Mellon Neuroscience Institute

Author Note

This work was supported by the National Science Foundation (NSF 1655126 to L. L. H. and NSF 1950054 to L.L.H and F.D.) and the National Institutes of Health (T32GM0081760). C.L.R. is now at the University of Pittsburgh, Department of Communication Science & Disorders. Stimuli and data are available at https://doi.org/10.17605/OSF.IO/9DKJG.

Abstract

Category learning is fundamental to cognition, but little is known about how it proceeds in real-world environments when learners do not have instructions to search for category-relevant information, do not make overt category decisions, and do not experience direct feedback. Prior research demonstrates that listeners can acquire task-irrelevant auditory categories incidentally as they engage in primarily visuomotor tasks. The current study examines the factors that support this incidental category learning. Three experiments systematically manipulated the relationship of four novel auditory categories with a consistent visual feature (color or location) that informed a simple behavioral keypress response regarding the visual feature. In both an in-person experiment and two online replications with extensions, incidental auditory category learning occurred reliably when category exemplars consistently aligned with visuomotor demands of the primary task, but not when they were misaligned. The presence of an additional irrelevant visual feature that was uncorrelated with the primary task demands neither enhanced nor harmed incidental learning. By contrast, incidental learning did not occur when auditory categories were aligned consistently with one visual feature, but the motor response in the primary task was aligned with another, category-unaligned visual feature. Moreover, category learning did not reliably occur across passive observation or when participants made a category-nonspecific, generic motor response. These findings show that incidental learning of categories is strongly mediated by the character of coincident behavior.

Keywords: category learning; incidental learning; statistical learning; auditory-visual associations

Categorization, the ability to treat distinct perceptual experiences as functionally equivalent, is a vital component of human cognition that underlies many everyday behaviors. Auditory categorization plays a role in deciphering words heard in a noisy restaurant, deciding quickly whether an approaching animal is friend or foe, and identifying one's own cell phone ring from that of others.

Although there is a rich literature on how humans learn categories, our understanding is largely based on laboratory studies conducted with visual objects and using training paradigms that involve explicit categorization. Typically, participants are aware that the objects should be sorted, make overt category decisions, and receive feedback that directs future decisions. This classic approach has provided an informative literature characterizing category learning (for reviews see Ashby & Maddox, 2011; Holt, 2011; Richler & Palmeri, 2014). Nonetheless, results obtained across overt training with visual objects may not generalize broadly to other perceptual modalities, or to natural environments that do not provide explicit training (Markman & Ross, 2003; Roark & Holt, 2018, 2019; Scharinger et al., 2013; Wade & Holt, 2005).

Indeed, category learning in natural environments typically occurs under much less explicit conditions. In everyday life, people do not usually receive instructions to search for category-diagnostic dimensions, make overt category decisions, or obtain explicit feedback. One possibility is that acquiring categories in the 'real world' involves accumulation of input regularities through purely passive exposure (Love, 2002; Maye et al., 2002; McMurray et al., 2009; Yoshida et al., 2010). Exposure to distributions of perceptual input in the lab is indeed known to affect perception of established speech categories (Clayards et al., 2008; Idemaru & Holt, 2011) and to influence judgments of a set of visual objects (Oriet & Hozempa, 2016). Likewise, listeners can learn novel nonspeech auditory categories experienced passively, when the exemplars comprising the categories exhibit coherent *a priori* similarity structure (Emberson et al., 2013; Wade & Holt, 2005).

However, there may be limits to the types of categories that can be acquired through mere exposure to auditory regularities. More complex nonspeech categories – like those modeling the multidimensional nature of speech categories – are not readily learned via passive exposure alone (Emberson et al., 2013; Wade & Holt, 2005). Infants are widely believed to form speech categories via passive exposure to individual exemplars (e.g., Maye, et al., 2002), but a recent meta-analysis of data available across 26 published studies (406 infants) concludes that infant category learning by passive exposure may be somewhat limited (Cristia, 2018). Yet, introducing alignment *across* modalities in passive observation may be beneficial for learning. Passive learning of statistical structure in the auditory modality can be improved by co-present, aligned visual cues (Cunillera et al., 2010; Mitchel & Weiss, 201; Thiessen, 2010). Thus, there is reason to believe that the alignment of statistical regularities in a learning environment will impact learning outcomes.

It is an open question as to how category learning proceeds when explicit feedback is unavailable, but passive accumulation of acoustic regularities may be sluggish or insufficient. In the auditory domain, motivated by speech category learning, there has been progress in developing new approaches to studying category learning under conditions that are neither explicit nor passive. In these studies, the approach has been to examine category learning in contexts in which participants interact actively with the category-relevant information in the context of an engaging primary task (Clapper, 2012; Gabay et al., 2015; Lim et al., 2019; Lim & Holt, 2011; Protopapas et al., 2017; Seitz et al., 2010; Vlahou et al., 2012; Wade & Holt, 2005). Notably, this primary task does not involve instruction about the existence of categories, overt category decisions, or explicit feedback about categorization. But, unbeknownst to participants primary task performance can be supported by category learning.

In such *incidental category learning* studies, sound categories are learned by virtue of their relationship to success in performing a task defined by other, and largely visuomotor, task demands. Although participants do not overtly search for dimensions diagnostic to auditory category membership and do not receive explicit feedback, incidental learning is neither passive, nor entirely unsupervised or feedback-free. There are rich associations of sound categories with behavioral events and outcomes in the primary task. Yet, unlike explicit category learning, the incidental feedback that might arise from these associations is not directly related to overt category decisions (Lim et al., 2019).

For example, in a task originally developed by Wade and Holt (2005; see also Kimball et al., 2013), the objective is to navigate a space-themed videogame, targeting approaching aliens with a laser. Participants are instructed only in how to maneuver in the game. They are not instructed 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 that includes a background score and acoustic events unrelated to the categories. 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 a 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* and each alien originates from a stereotypical region of visual space. Thus, learning to categorize across the acoustically-variable sounds associated with specific aliens supports faster action. Indeed, participants quickly learn both novel artificial nonspeech auditory categories (Lim et al., 2015, 2019; Wade & Holt, 2005) and also non-native speech categories (Lim et al., 2015; Lim & Holt, 2011; Wiener et al., 2019). Moreover, they generalize this learning to novel category exemplars in a post-game overt labeling task in which novel sounds are matched with alien creatures.

Although other-worldly, this task's demands more closely approximate those of learning in a natural environment than traditional explicit learning or passive-exposure paradigms. Here, sound categories are not themselves the focus of the task, but instead convey information vital for recognizing events taking appropriate actions in order to 'survive and thrive'. Such paradigms capture some of the incidental nature of learning categories in everyday life.

Nonetheless, these games impose a trade-off: because they are necessarily more complex and dynamic, it is difficult to uncover which of the videogame's many elements are responsible for driving learning. To address this issue, Gabay et al. (2015) developed a task that, while much simpler, targets aspects of the videogame hypothesized to drive incidental learning. In this Systematic Multimodal Association Reaction Time (SMART) 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. Unbeknownst to participants, and like the Wade and Holt (2005) videogame, sounds are drawn from one of four distinct sound categories. There is a multimodal (auditory-category to visual-location) correspondence that relates variable sound category exemplars to the location where a visual object will consistently appear. Again, like the videogame, 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, although in no way *necessary* for task completion.

In the training blocks, 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 possibly even awareness – of the existence of categories. Participants are never told 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 via a ‘random’ block of trials. Here, the relationship between auditory category and visual location is eliminated. Thus, if participants incidentally learn sound categories to support quick detection of the visual target, then in the 'random' block, visual detection should be slower relative to the previous training block for which audiovisual relationships were coherent – in other words, a reaction time cost (RT Cost).

Additionally, an overt sound categorization post-test follows the SMART task. In this task, 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 overt or incidental feedback. This provides a measure of generalization of learning to novel exemplars, a hallmark of robust category learning. It also requires that participants transfer learning to an explicit task that differs somewhat from the incidental SMART learning context. It is impossible to succeed at generalizing category knowledge in this explicit labeling task without having learned incidentally in the SMART task. As this test involves only *novel* category exemplars that were not experienced in the incidental learning context of the SMART task, it is impossible to perform this task above chance level by memorizing training exemplars. Further, the original Gabay et al. (2015) paper introducing the SMART task included additional control experiments that support this conclusion.

The SMART task shares some characteristics with a traditional procedural learning paradigm, the serial reaction time (SRT) task (Nissen & Bullemer, 1987). However, SMART – at least as described above – measures category learning, and ***not*** sequence learning. There are no embedded sequences across the trials, which instead are randomly ordered across locations. SRT and SMART do share the fact that participants are not alerted to a regularity in the training stimuli. This is done to promote incidental learning conditions, which are distinct from both overt and passive paradigms. Here, we make no claims about the extent to which SMART is implicit; in fact, to foreshadow, it appears different samples of participants tested in different contexts may engage in quite different strategies during incidental learning.

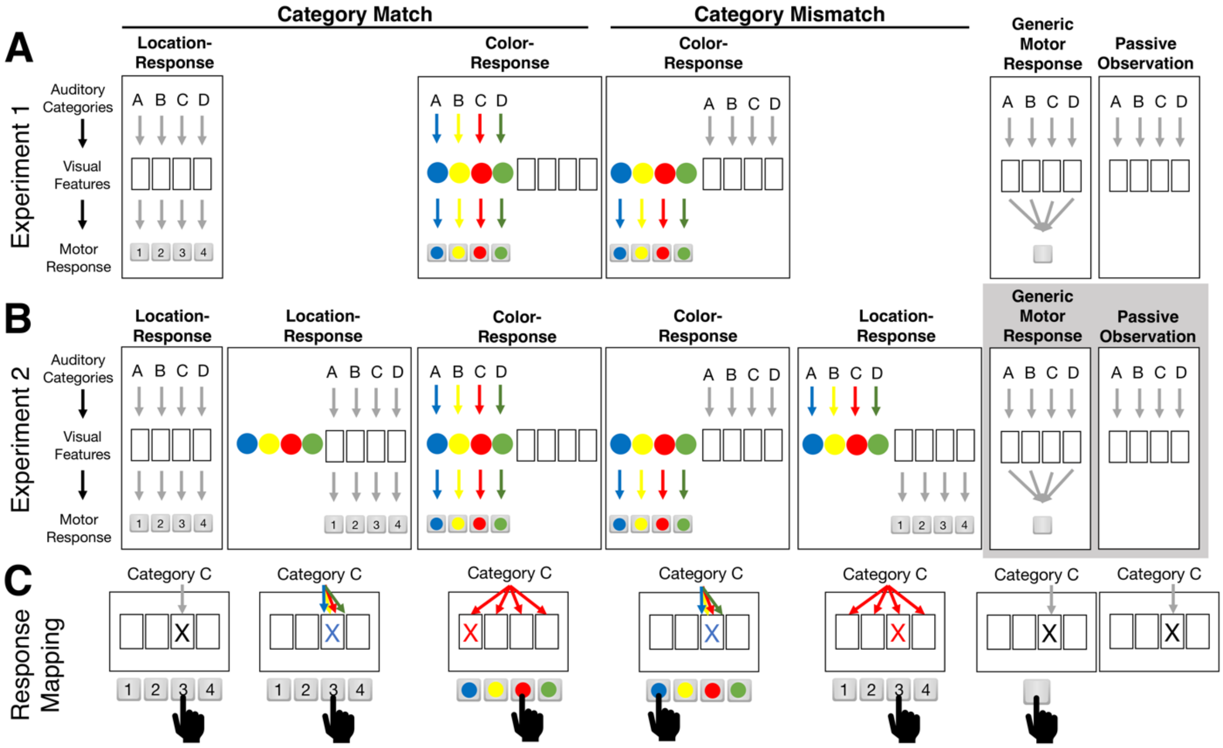
Using the SMART paradigm, Gabay et al. (2015) examined nonspeech auditory category learning across the same sound exemplars employed by Wade and Holt (2005) in the incidental videogame paradigm. Although the task was a simple visual detection, participants nonetheless learned the auditory categories. This was demonstrated first by the longer RTs to detect the visual target in the 'random' block, when the association between the sound categories and the upcoming location of the visual target was destroyed. The RT Cost imposed by the loss of systematic mapping between sound category and target location suggested that participants relied on incidentally learned auditory category information to facilitate speedy visual target detection. Moreover, this learning generalized to labeling novel auditory exemplars in the post-training test.

The SMART paradigm is a potentially attractive means of addressing a number of open questions about what factors drive – or, alternatively, hinder – incidental auditory category learning. Across three experiments with seven unique conditions (with five of these replicated across experiments), we systematically manipulate the relationship of four novel auditory categories with a consistent visual feature (color or location) that informs a simple behavioral keypress response regarding the visual feature in the SMART task. We first replicate the basic result reported by Gabay et al. (2015). We next ask whether incidental learning is dependent on the consistent *alignment* of auditory categories with the visuomotor demands of the primary task. We examine how the presence of an additional visual feature in the primary task that is uncorrelated with auditory categories or motor response impacts incidental learning. Finally, we test whether learning of the auditory categories would occur when participants experience the same auditory-visual statistical regularities via passive observation or with a generic motor response to the visual target that does not differentiate categories.

The three experiments are differentiated largely by their approach. Experiment 1 uses traditional in-laboratory examination of university students in a carefully controlled environment with studies conducted across an extended period. Experiments 2 and 3 replicate and extend Experiment 1 with online testing of a more diverse participant pool with random assignment to conditions over a brief period.

Figure 1

Condition Design Overview



Note. **A.** Design of Experiment 1 conditions in terms of the relationship between auditory categories (top row), visual features of spatial location and color (middle row) and motor response (bottom row). **B.** Design of Experiment 2 conditions with the same organization as **A**. Experiment 3 conditions are outlined in the gray box. **C.** Example of a single trial for a Category C exemplar across all conditions. **Category Match Conditions** establish a deterministic relationship between auditory categories and one visual feature of the target (color *or* location), which is also aligned with the unique motor response. *Category Match: Location-Response (no color)* (far left) is a near replication of the SMART task paradigm from Gabay et al. (2015) with a consistent relationship between each auditory category and the location of a visual target 'X'. Response is aligned with location (Location-Response). In *Category Match: Color-Response* and *Category Match: Location-Response* conditions, there is a consistent relationship between the categories and one of the visual features; the other feature irrelevantly varies. **Category Mismatch** conditions break the alignment between the auditory category-relevant visual feature and the motor-relevant feature. The **Generic Motor Response** condition maintains a consistent relationship between auditory categories and the visual location of the target, but participants make a single motor response to all stimuli (spacebar). The **Passive Observation** condition requires no motor response while participants observe a consistent auditory category to visual location mapping.

Experiment 1

Method

Experiment 1 involved five conditions manipulating characteristics of the SMART task to understand the factors supporting incidental category learning (Figure 1A). There are independent groups of learners participating in each condition, and learning is examined within as well as between conditions. Two major components differ across the five conditions: (1) the alignment of the auditory categories and task-relevant visual features of the primary visuomotor task, and (2) the motor response.

In the 'baseline' *Category Match: Location-Response (no color)* condition*,* each auditory category corresponds with a single greyscale rectangle location on the screen, and the X in a location guides a unique motor response. Thus, auditory category, location, and motor response are all aligned. This situation reflects a *match* between categories and the behaviorally relevant visual feature. This condition is a near replication of the SMART task developed by Gabay et al. (2015).

To assess the importance of the *differential alignment* of auditory categories with one of two behaviorally relevant visual features, we devised two new conditions. In *Category Match: Color-Response,* auditory categories are aligned with a different visual feature, color, which directs the motor response. In this condition, location is not associated with either auditory categories or motor response. *Category Mismatch: Color-Response* introduces a misalignment between auditory categories and the visual feature that guides motor response. Here, auditory category is associated with the *visual target location*, as in the first condition, but the motor response is based on color, which is uncorrelated with auditory category. This situation reflects a *mismatch* between categories and the behaviorally relevant visual feature.

To assess the importance of category-specific versus category-general *motor response* during incidental learning, we devised the *Generic Motor Response* condition. Although auditory categories are again uniquely linked to the visual target locations, participants push the spacebar when a target appears, regardless of its location. Finally, to assess the importance of explicit motor involvement, the *Passive Observation* condition removes the motor task completely, with participants engaging in passive observation of the auditory category to visual location association.

A power analysis (calculated using G*Power and the *pwr* package in R) using the effect size from comparison of the performance of two groups in the generalization test from Gabay et al. (2015, Experiment 1 vs. 3) indicated that to detect a large difference among conditions ($d = 0.87$ or $f = 0.4$), a sample of 21 participants per group would be needed to obtain statistical power at a .90 level with an alpha of .05. Thus, for each condition, we met or exceeded this target of 21 in recruitment and tested 120 participants across conditions. The protocols for all experiments were approved by the Institutional Review Board at Carnegie Mellon University. Data for all experiments are available at https://doi.org/10.17605/OSF.IO/9DKJG.

Participants

Across conditions, participants were 120 (80 F, 40 M) members of the Carnegie Mellon University community ages 18-29 years who were either given partial course credit or $10 for participating. There were 21 (15 F, 6 M) participants in *Category Match: Location-Response (no color)*; 21 (13 F, 8 M) in *Category Match: Color-Response*; 24 (14 F, 10 M) in *Category Mismatch: Color-Response*; 23 (16 F, 7 M) in *Generic Motor Response*; and 31 (22 F, 9 M) in *Passive Observation*. Two *Passive Observation* participants were withheld from analyses: one due to poor performance on catch trials (see below) and one due to failure to complete the entire task, leaving 29 participants. There are slightly more participants in this condition because we wanted to ensure results robust to potential individual variability in learning from passive observation not guided by response engagement. All participants had normal or corrected-to-normal vision and reported normal hearing.

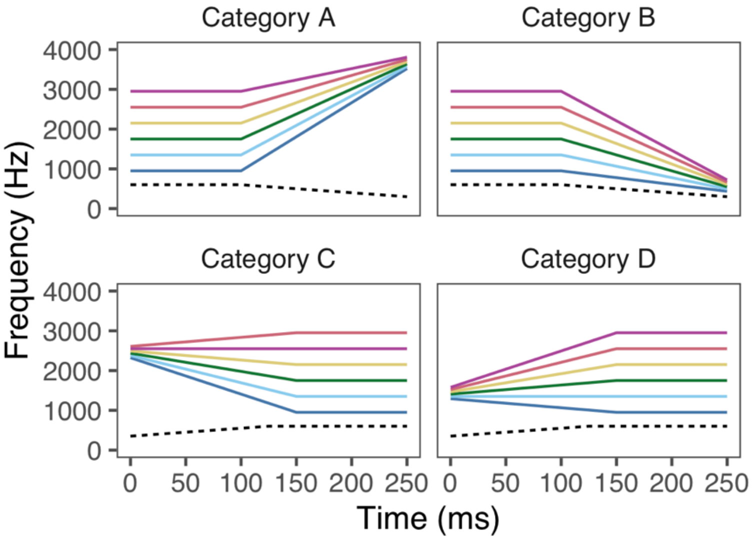
We note that Carnegie Mellon University community participants were *not* randomly assigned to condition in Experiment 1, presenting a potential limitation. Experiments 2 and 3 address this concern in replicating and clarifying results of Experiment 1, with random assignment in independent online samples and data collection accomplished in a brief (3-week, Experiment 2; 1-day, Experiment 3) period.

Stimuli

For all conditions, the auditory categories were defined by novel 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. The stimuli are available at https://doi.org/10.17605/OSF.IO/9DKJG. 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 five novel exemplars per category 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 2, categories A and B). The other two categories were defined in a more complex, higher-dimensional perceptual space (no single acoustic cue uniquely defined category membership, Figure 2, categories C and D). Each exemplar was 250 ms in duration and exemplars were matched in RMS amplitude.

Figure 2

Auditory Categories



Note. Auditory categories. Each higher-frequency component (colored lines) is paired with the lower-frequency component (dashed line) to create six exemplars for each category presented during training. The five generalization exemplars in each category are not shown. The stimuli are available at https://doi.org/10.17605/OSF.IO/9DKJG and described in more detail in Wade and Holt (2005)

Procedure

SMART task. All conditions make some variation of the SMART task from Gabay et al. (2015; see Figure 1). We first describe the procedure for the 'baseline' *Category Match: Location-Response (no color)* condition, which is a near replication of the SMART task developed by Gabay et al. (2015). We then describe the key differences in the methods for the other conditions that manipulate this basic method.

In the SMART task, participants saw four greyscale boxes on the screen aligned horizontally in a straight line that corresponded with the participant response keys ('*u*', '*i*', '*o*', '*p*'). On each trial, participants heard five unique 250-ms sound exemplars drawn from one of the four auditory categories (50 ms ISI, 1500 ms total duration). Immediately following the final auditory stimulus, participants saw an X appear in one of the four boxes. Participants in *Category Match: Location-Response (no color)* were instructed to report the location of the X with a unique button press associated with each location. Reaction time was measured from the onset of the visual target to the keypress.

Participants completed 5 blocks of the SMART task (384 total trials). In Blocks 1-3 and 5, the sound category perfectly predicted the location of the upcoming visual target across 96 trials. In Block 4, the relationship between the sound category and location of the X was random, such that after five different exemplars from a single category played, the X appeared with equal likelihood in any of the four boxes. In Block 4, there were only 48 trials (half as many as the other blocks) to limit the impact of this randomization on the ultimate learning outcomes. Following the approach of Gabay et al. (2015) study, we examined the reaction time (RT) in Block 4 compared to Block 3 (when category-to-location association was destroyed) to evaluate the cost of randomizing the association of a sequence of five coherent auditory category exemplars with the location of the upcoming visual target. We refer to this difference as the RT Cost, which serves as a covert index of incidental auditory category learning.

One small change was made to the Gabay et al. (2015) paradigm. Gabay et al. (2015) randomized Block 4 trials such that each of the five auditory stimuli could be drawn from any auditory category. Therefore, not only did the sounds convey no information about target location, there was also no consistent category membership across the five exemplars played in a trial. By contrast, the current study maintains within-category coherence within Block 4 trials, which still enables participants to use the auditory information to predict where the target will appear and prepare a motor response. In all blocks, the sounds on a single trial are drawn from the same auditory category, but in Block 4 the category-to-location mapping is completely randomized.

Generalization test. After completing the SMART task, participants were informed that there was a relationship between the sounds they had heard, and the location of the visual target. They were then asked to complete a 4-alternative forced choice (4AFC) task with no feedback. In each of 96 trials, participants heard five unique exemplars from one category, and were asked to select which location they believed a target might appear. Each of these exemplars were novel and had not appeared in the SMART task; thus, accurate response required generalization of category learning.

Key differences across conditions. The other conditions make small variations to the 'baseline' *Category Match: Location-Response (no color)* condition SMART task to understand the factors supporting successful learning. Across each condition, the stimuli and basic procedure were the same. Below, key manipulations are described for each condition.

Category Match: Color-Response. In the 'baseline' condition, sound categories predict the spatial location of the visual target. In contrast, in the *Category Match: Color-Response* condition, sound categories predict the *color* of the visual target. Instead of a single black visual target ('X'), in *Category Match: Color-Response* there are four visual targets distinguished by color (red, blue, green, and yellow 'X' targets). Each of the distinctly colored targets appeared in each location with equal probability, such that visual target location was completely unassociated with the auditory categories. Crucially, *participants responded based on the color of the target and not the location*. Colored stickers affixed to response keys facilitated the color-response mapping (Figure 1). During the generalization test, participants were informed of the relationship between the sounds and color of the visual target and completed a 4AFC task with no feedback.

Category Mismatch: Color-Response. The *Category Mismatch: Color-Response* condition was almost identical to *Category Match: Color-Response*, with one key difference: the *auditory categories* were associated with the target *location*, but the *motor response* was based on target *color*. Thus, there was a misalignment between the auditory categories and the behaviorally relevant visual feature. The auditory category predicted the *location* of the target X, with distinctly colored targets appearing in each location with equal probability. After the SMART task, participants were informed of the relationship between the sounds and location of the visual target and completed the 4AFC generalization task with no feedback.

This manipulation decoupled the association between the category-associated visual feature and the task-relevant visual feature. If a coherent category-visual association is sufficient to drive incidental category learning, then we expect to observe learning. However, if the alignment of auditory categories with a task- and motor-relevant visual feature is necessary for learning, then decoupling the auditory categories from the response should eliminate or reduce incidental category learning.

Generic Motor Response. The procedure for the *Generic Motor Response* condition was also nearly identical to the baseline: auditory categories mapped to visual target location and all targets were the same color. But unlike the baseline condition, participants pressed only the spacebar to indicate detection of a visual target instead of reporting its location. At the start of the generalization test, participants were instructed about the spatial mapping of the visual target locations and the '*u', 'i', 'o', 'p'* keys on the keyboard, and made responses based on this mapping.

If incidental learning of the auditory categories requires a response unique to each category, we predict no category learning effects in *Generic Motor Response* on either the covert RT Cost measure or the overt generalization test. If instead, consistent auditory category mapping to visual target location plus the mere execution of a timed, yet generic, motor response is sufficient to support incidental learning, then we should observe learning.

Passive Observation. In the *Passive Observation* condition, the procedure was similar to the baseline *Category Match: Location-Response (no color)*. Again, auditory categories were linked to spatial location, but participants made no motor responses during training. This manipulation is critical. The lack of a motor response makes this task entirely passive, but participants observe the auditory-category to visual-location correspondence. As a result, this paradigm can be distinguished both from the incidental learning paradigms of each of the other conditions (which rely on active engagement in the visuomotor task), as well as from completely unsupervised auditory category learning, since the association of the auditory categories and visual location can serve as a form of feedback (similar to Cunillera et al., 2010; Mitchel & Weiss, 2011; Thiessen, 2010).

The *Passive Observation* condition included some additional components to ensure that the experience of participants was as similar as possible to that of other conditions, even in the absence of a trial-wise motor response. First, following the last of the five category exemplars presented on a given trial, the visual target 'X' remained on the screen for 500 ms. This value was based on the average RT across all subjects and all blocks for *Category Match: Location-Response (no color)* (M = 432 ms), rounded up to accommodate modest individual differences. Second, we included several catch trials in each block to assure participants remained engaged with the audio-visual stimuli, despite the lack of trial-by-trial responses. Catch trials were signaled in the following way: after visual targets appeared on the screen in the typical manner, they subsequently appeared across all four visual locations (maximum 2 seconds). Catch trials comprised 12.5% of the total trials: 12 trials for every 96-trial block, 6 trials in Block 4, randomly dispersed. Participants were instructed to quickly press the spacebar when they saw flashing Xs. If participants did not respond within 2 seconds, they received feedback to respond faster.

Participants responded within the 2 sec required on 99.5% of trials. We planned to exclude any participants who did not respond or responded too slowly on more than 5% of catch trials. As a result of this criterion, one participant (94% accurate) was excluded from analyses. Including this participant did not change any of the results.

Together, the *Generic Motor Response* and *Passive Observation* conditions allow us to test hypotheses about how the motor response supports learning in the SMART task. If *any* motor response linked with the visual target supports learning, we should observe learning in *Generic Motor Response*, but not *Passive Observation*. If motor response is not required for learning, we should observe learning in each of these conditions. Finally, if a unique motor response to visual target is required, we should observe a lack of learning in these conditions.

Summary of Conditions in Experiment 1

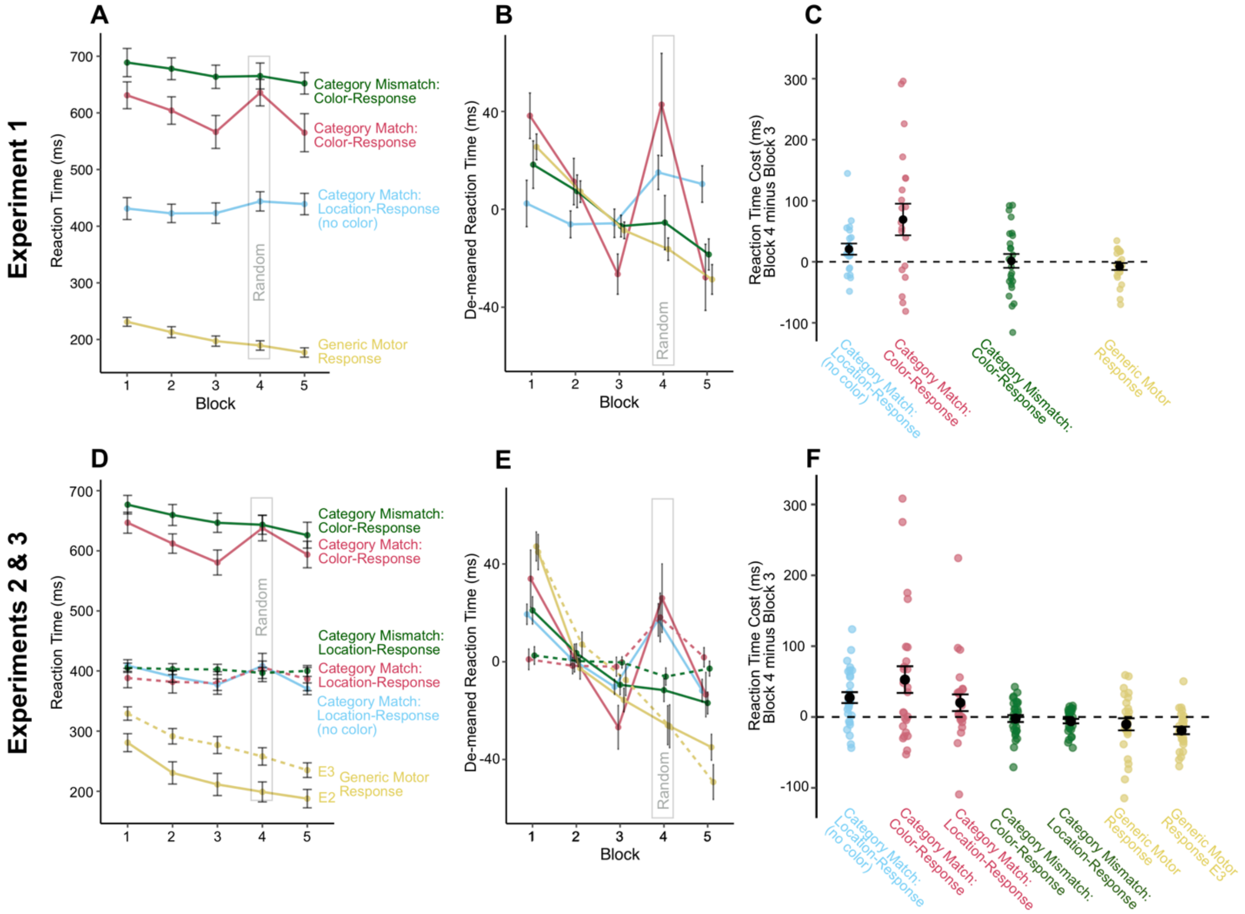
Category Match: Location-Response (no color) is a near replication of the SMART task from Gabay et al. (2015) and forms the basis of comparison across Experiment 1 manipulations. *Category Match: Color-Response* switches the category-and-task-relevant feature from the location of the 'X' on the screen to a new visual feature – color – but retains trial-to-trial variability in target location, which is task- and category-irrelevant. Like *Category Match: Color-Response*, *Category Mismatch: Color-Response* maintains one task-relevant and one task-irrelevant visual feature. However, in *Category Mismatch: Color-Response*, auditory categories are associated with the *task-irrelevant* visual feature (spatial location, as in the baseline), thereby breaking the association between auditory categories and motor response. In *Generic Motor Response*, we remove the uniqueness of this motor response: instead, participants make the same keypress (spacebar) on each trial. This allows for examination of whether incidental learning depends upon a unique category-to-motor-response mapping. Finally, in *Passive Observation*, we completely remove motor responses to determine whether passive observation of a coherent auditory category-visual feature association is sufficient to drive category learning. Figure 1 summarizes the conditions.

Results

To compare learning, we examined the reaction times to detect the visual target from visual target onset, the online measure of learning (RT Cost from Block 3 to Block 4), and the overt generalization post-test accuracy. Here and throughout, we used Welch's one-way ANOVA, Welch's *t*-tests and Games-Howell post-hoc tests to compare across conditions, which do not assume homogeneity of variances. Other model assumptions were met.

Figure 3

Reaction Time Measures Across Experiments



Note. Results for Experiment 1 are shown on the top (A-C) and Experiments 2 and 3 are shown on the bottom (D-F). **A/D.** Average, unnormalized reaction times across all subjects for each condition except for the Passive Observation condition. In Experiment 2, conditions that are direct replications of Experiment 1 are represented with solid lines and extension conditions are represented with dashed lines. The *Generic Motor Response* and *Passive Observation* conditions of Experiment 3 are shown alongside Experiment 2 as dashed lines. **B/E.** Average, participant-wise de-meaned reaction time across the five blocks. **C/F.** Individual subject and average reaction time costs for each condition. The dashed line at 0 reflects no Block 4 and Block 3 reaction time difference (no RT Cost). Positive values reflect a slowing in visual target detection when the category-to-target association is destroyed. Error bars reflect the standard error of the mean (*SEM*). Spacing of the conditions along the x-axis reflects the alignment of the replication conditions across the two experiments.

Reaction Time Measures

Pre-processing. Because the *Passive Observation* condition did not require trial-by-trial responses, we examined RT measures only for the other four conditions. On average, participants were more than 95% accurate in visual detection in SMART training across all conditions. As in Gabay et al. (2015), trials were excluded from analyses for which there was a visual detection error, or reaction time was less than 150 ms or greater than 1500 ms; these boundaries were greater than two standard deviations calculated from the most variable condition: *Category Misaligned: Color-Response*. This led to a total of 3.8% of trials removed for *Category Match: Location-Response (no color)*; 7.9% of trials for *Category Match: Color-Response*; and 7.7% of trials for the *Category Mismatch: Color-Response*. Because the *Generic Motor Response* condition required only visual detection rather than mapping to a response, we expanded the lower bound of this criterion and excluded trials with reaction times less than 50 ms or greater than 1500 ms (6.8% of trials). We removed trials from analyses based on an *a priori* data analysis plan aligned with the approach of Gabay et al. (2015). However, we note that the results do not change when all trials are included in the analyses.

De-meaned Reaction Times. Because task demands across conditions were slightly different, there are some differences in overall RTs that make a direct comparison of average RT differences difficult to interpret (Figure 3A). These relative differences in overall RT could indicate differences about the effort or demand of each individual task, with higher RTs reflective of more challenging tasks (e.g., mapping response to color versus location). To visualize the relative differences in RT across conditions while controlling for these overall differences, we computed the de-meaned reaction times for each of the conditions (Figure 3B). Specifically, for each participant, the mean RT across the non-random training blocks (1-3, and 5) was used as a baseline. Then, the RT for each trial for each participant was normalized using this baseline measure ($RT_{normalized}$ = $RT_{unnormalized}$ - RT_{mean}). This visualization illustrates the relative change in RT across blocks, with an increase in RT in Block 4 (RT Cost) for *Category Match: Location-Response (no color)* and *Category Match: Color-Response* conditions.

Reaction Time Cost. Incidental learning within the SMART paradigm is measured as the RT Cost to visual detection from Block 3 to Block 4 (Figure 3C), for which the relationship between auditory categories and visual features is eliminated through randomizing category-to-feature assignment.

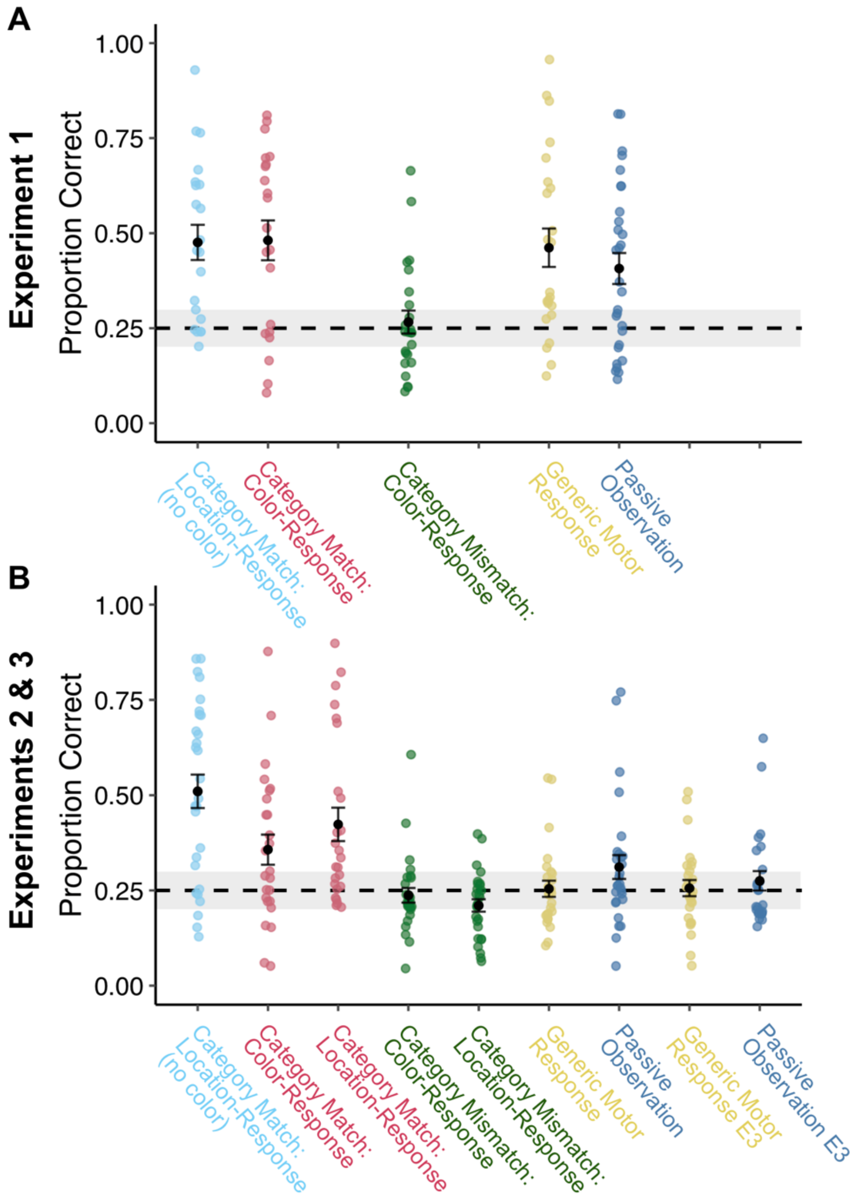
We first examined whether there was a significant RT Cost for each condition with paired samples *t*-tests across Block 3 and Block 4 RTs. *Category Match: Location-Response (no color)*, participants had a significant RT Cost with a mean of 20.8 ms ($t(20) = 2.28$, $p = .034$, $d = 0.50$, 95% CI [1.74, 39.8]). *Category Match: Color-Response*, participants had a significant RT Cost with a mean of 69.3 ms ($t(20) = 2.68$, $p = .015$, $d = 0.58$, 95% CI [15.3, 123.3]). By contrast, there was no significant RT Cost in *Category Mismatch: Color-Response* ($M = 1.43$ ms, $t(23) = 0.13$, $p = .90$, $d = 0.03$, 95% CI [-22.06, 24.9]) or *Generic Motor Response* conditions ($M = -7.67$ ms; $t(22) = 1.34$, $p = .19$, $d = 0.28$, 95% CI [-19.5, 4.16]).

Next, we asked whether the magnitude of the RT Costs differed across conditions. A between-subjects Welch's ANOVA indicated they did (Welch's $F(3, 43.2) = 4.48$, $p = .0078$, est. $\omega^2 = 0.10$). We compared RT Costs across conditions to assess the influence of the *alignment* of the auditory-visual-motor information and the *motor response* on RT Cost. Whereas there were differences in the presence or absence of an RT Cost across conditions, these were not significantly different when comparing across conditions that differ in the alignment (*Category Match: Location-Response (no color)*, *Category Match: Color-Response, Category Mismatch: Color-*

*Response, p*s > .10) or motor response demands (*Category Match: Location-Response (No Color)* vs. *Generic Motor Response, p* = .057).

Figure 4

Generalization Test Accuracy Across All Conditions



Note. Accuracy averaged across all four categories relative to chance (25%, dashed line with 95% CI across 96 trials in shaded region). Error bars reflect *SEM*. **A.** Experiment 1 **B.** Experiments 2 and 3. Spacing of the conditions along the x-axis reflects the alignment of the replication conditions across experiments.

Generalization Test Accuracy

The same overt generalization post-test was used to examine category learning in each of the five conditions (Figure 4A). First, we examined accuracy in overtly mapping the auditory category-to-response association established during the training phase in the SMART paradigm, relative to chance performance (25%). In 'baseline' *Category Match: Location-Response (no color)*, participants successfully generalized their knowledge, with an average accuracy of 47.6%, significantly greater than chance ($t(20) = 4.89$, $p < .001$, $d = 1.07$, 95% CI [37.9, 57.2]). In *Category Match: Color-Response*, where the task-relevant feature (color) was predicted by auditory category and visual target location varied randomly, performance was above chance ($M = 48.1\%$, $t(20) = 4.42$, $p < .001$, $d = 0.96$, 95% CI [37.2, 59.0]). In *Category Mismatch: Color-Response*, where auditory category was *not* linked to the task-relevant feature (color), but instead aligned with the task-irrelevant feature (location), there was no evidence of learning ($M = 26.6\%$; $t(23) = 0.53, p = .60, d = 0.11$, 95% CI [20.4, 32.9]). By contrast, in *Generic Motor Response* ($M = 46.2\%$, $t(22) = 4.19, p = .00038, d = 0.87$, 95% CI [35.7, 56.6]) and *Passive Observation* ($M = 40.7\%$, $t(28) = 3.86, p = .00061, d = 0.72$, 95% CI [32.4, 49.0]) performance was above-chance.

Post-test accuracy varied across conditions (Welch's $F(4, 54.4) = 6.1$, $p = .00040$, est. $\omega^2 = 0.15$), such that generalization of category learning at post-test was associated with alignment of auditory category, visual feature, and motor response. Accuracy was significantly higher when category-response aligned in the *Category Match: Location-Response (no color)* and *Category Match: Color-Response* conditions compared to *Category Mismatch: Color-Response* for which category-response was misaligned ($p = .005$ and 0.01, respectively); Category Match conditions did not differ ($p = 1.0$).

Involvement of a category-unique versus generic versus no motor response did not impact expression of incidental learning in the overt task. Post-test accuracy for baseline *Category Match: Location-Response (no color)* did not differ significantly from *Generic Motor Response* (*p* = 1.0) or *Passive Observation* conditions (*p* = .80), which themselves did not differ (*p* = .92).

Discussion

Together, these findings demonstrate the robustness of incidental learning in the SMART paradigm and indicate that a key factor driving learning is the alignment between auditory categories and behaviorally relevant features of the primary task, here visual features of location or color. Successful overt generalization occurs when auditory categories are aligned with a task-relevant visual feature (*Category Match: Location-Response (no color)*), even if there is a task-irrelevant visual feature (*Category Match: Color-Response*). Learning is hindered when auditory categories are no longer associated with the task-relevant visual feature (*Category Mismatch: Color-Response*), even when there is the same valid auditory-visual association as in *Category Match: Color-Response*. Seemingly in contrast to these findings of the importance of category-coupled motor responses, participants in the *Generic Motor Response* and *Passive Observation* conditions also showed significant auditory category learning in the explicit categorization post-test.

Experiment 2

To establish the robustness of these findings, particularly in view of the potential for cohort effects due to the lack of random assignment in Experiment 1, we conducted a replication and extension experiment. Here, we tested the same five conditions as in Experiment 1 and added two additional conditions that serve as counterbalanced versions of the *Category Match: Color-Response* and *Category Mismatch: Color-Response* conditions to examine how the particular visual feature associated with the auditory categories might impact learning. In addition, we move from the carefully controlled laboratory environment of Experiment 1 to online testing, with a more heterogeneous sample of participants and random assignment to condition.

Methods

Participants

Participants were recruited via Prolific (www.prolific.sc) and participated via the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2019) across a 3-week period in April-May 2021. Participants were randomly assigned to one of seven conditions. Across conditions, there were 209 participants (55 F, 150 M, 3 non-binary, 1 'prefer not to answer') ages 18-35 years who were paid $10/hour for participating. By condition, there were: 30 (7 F, 22 M, 1 'prefer not to answer') participants in 'baseline' *Category Match: Location-Response (no color)*; 30 (4 F, 26 M) in *Category Match: Color-Response*; 29 (8 F, 21 M) in *Category Match: Location-Response*; 29 (12 F, 16 M, 1 non-binary) in *Category Mismatch: Color-Response*; 30 (9 F, 21 M) in *Category Mismatch: Location-Response*; 31 (8 F, 22 M, 1 non-binary) in *Generic Motor Response*; and 30 (7 F, 22 M, 1 non-binary) in *Passive Observation*.

Because this study was conducted online, we included additional measures to ensure task compliance. Specifically, we used a headphone check task that utilizes dichotic pitch (Milne et al., 2020) and introduced additional catch trials in each condition. Participants were required to pass the headphone check prior to beginning the experiment. Due to poor performance on catch trials (see below), a total of 21 participants were withheld from analyses: 2 from *Category Match: Location-Response (no color)*; 5 from *Category Match: Color-Response*; 4 from *Category Match: Location-Response*; 1 from *Category Mismatch: Color-Response*; 2 from *Category Mismatch:*

Location-Response; 5 from *Generic Motor Response*; and 2 from *Passive Observation*. All participants reported normal or corrected-to-normal vision and normal hearing.

Procedure

The stimuli and general procedure were the same as Experiment 1. As noted above, in addition to replicating the five conditions from Experiment 1, we added two additional conditions, summarized in Figure 1B. These conditions simply change which visual feature is aligned or misaligned with the auditory categories. The *Category Match: Location-Response* condition aligns the auditory categories and response with the location of the target while color irrelevantly varies. The *Category Mismatch: Location-Response* condition aligns auditory categories with color, but response is based on location.

Different from Experiment 1, participants pressed the '*d*', '*f*', '*j*', and '*k*' keys instead of '*u*', '*i*', '*o*', and '*p*'. This served to control for potential differences in which finger was used to make responses, ensuring participants used more dominant fingers to respond. We also included attention catch trials across all conditions, described below. Because we could not place color maps on online participants' keyboards, the color mapping was present on the screen at all times during the task for conditions requiring a color response.

Catch trials. In *Generic Motor Response* and *Passive Observation* conditions, we included auditory and visual catch trials to ensure that participants attended to auditory *and* visual information. In all other conditions, we included auditory catch trials only, as we were able to measure how well participants were attending to the visual information based on their SMART task performance. There were 12 auditory catch trials and – where applicable – 12 visual catch trials for every 96-trial block (6 trials in Block 4), all randomly dispersed. Visual catch trials were identical to Experiment 1. Auditory catch trials were similar, except that participants were instructed to quickly press the spacebar (or the '*t*' key in *Generic Motor Response* condition) as soon as they heard a 200 ms 1000 Hz pure tone. If participants did not respond within 2 seconds, they received feedback to respond faster.

We excluded participants who did not respond within the required time on more than 15% of trials. As a result of this criterion, 21 participants across all seven conditions were excluded from analyses. Included participants responded correctly on 97% of catch trials; excluded participants responded correctly on 39% of catch trials. We included this more liberal criterion relative to Experiment 1 to accommodate for modest differences in online participants compared to in-laboratory participants. We note that using a 15% criterion to exclude participants in Experiment 1 does not change the Experiment 1 results.

Results

We first present the results from the replication conditions addressing the questions about *alignment* and *motor response* introduced in Experiment 1. We then present the results from the new Experiment 2 conditions to understand how the visual feature associated with the auditory categories influenced results. A comparison of the RT Cost and generalization test accuracies across experiments can be found in Table 1.

Table 1

Comparison of Results in All Experiments

	Experiment 1		Experiment 2		Experiment 3	
	SMART RT Cost	Post-Test Accuracy	SMART RT Cost	Post-Test Accuracy	SMART RT Cost	Post-Test Accuracy
Category Match: Location-Response (no color)	20.8 ms	47.6%	27.4 ms	51.0%	-	-
Category Match: Color-Response	69.3 ms	48.1%	52.8 ms	35.7%	-	-
Category Match: Location-Response	-	-	20.1 ms	42.3%	-	-
Category Mismatch: Color-Response	1.43 ms	26.6%	-2.18 ms	23.7%	-	-
Category Mismatch: Location-Response	-	-	-5.80 ms	21.1%	-	-
Generic Motor Response	-7.67 ms	46.2%	-10.3 ms	25.4%	-19.0 ms	25.6%
Passive Observation	-	40.7%	-	31.1%	-	27.5%

Reaction Time Measures

Pre-processing. As in Experiment 1, visual target detection in SMART training was highly accurate across all conditions (> 91%). Trials for which there was a visual detection error or reaction time was less than 150 ms or greater than 1500 ms were excluded from analyses. For the *Generic Motor Response* condition, reaction times less than 50 ms were excluded. This resulted in the following exclusions: 4.5% *Category Match: Location-Response (no color)*; 9.8% *Category Match: Color-Response;* 6.9% *Category Match: Location-Response;* 7.0% *Category Mismatch: Color-Response*; 4.9% *Category Mismatch: Location-Response*; 8.7% *Generic Motor Response.*

As in Experiment 1, data exclusions were planned *a priori* according to prior research and including all trials, without exclusions, did not change the outcomes of any analyses.

De-meaned Reaction Times. As in Experiment 1, we plotted the average RTs (Figure 3D) and the de-meaned RTs across blocks for all conditions (Figure 3E).

Reaction Time Cost. RT Cost, the difference in RT between Block 3 and Block 4, is the main measure of interest from SMART training (Figure 3F). As in Experiment 1, RT Cost estimates the extent to which the task-irrelevant relationship between auditory categories and visual features established in Blocks 1-3 impacts behavior on the task-relevant response to the visual target.

As in Experiment 1, paired-samples *t*-tests of RT across Block 3 and Block 4 revealed a significant RT Cost for baseline *Category Match: Location-Response (no color)* (*M* = 27.4 ms; *t*(27) = 3.55, *p* = .0014, *d* = 0.67, 95% CI [11.6, 43.2]) and *Category Match: Color-Response* (*M* = 52.8 ms; *t*(24) = 2.80, *p* = .0098, *d* = 0.56, 95% CI [13.9, 91.7]). The new Experiment 2 condition *Category Match: Location-Response* was also positive but not significant (*M* = 20.1 ms, *t*(24) = 1.71, *p* = .10, *d* = 0.34, 95% CI [-4.18, 44.5]). There was no significant RT Cost for *Category Mismatch: Color-Response* (auditory-location, color-response; *M* = -2.18 ms; *t*(27) = -0.45, *p* = .66, *d* = 0.084, 95% CI [-12.2, 7.87]), *Category Mismatch: Location-Response* (auditory-color location-response; *M* = -5.80 ms; *t*(27) = -1.85, *p* = .075, *d* = 0.35, 95% CI [-12.2, 0.63]), or *Generic Motor Response* conditions (*M* = -10.3 ms; *t*(25) = -1.21, *p* = .24, *d* = 0.24, 95% CI [-27.9, 7.25]).

We next asked whether the magnitude of RT Costs differed across the conditions in Experiment 1 replicated in Experiment 2 (*Category Match: Location-Response (no color)*, *Category Match: Color-Response*, *Category Mismatch: Color-Response*, and *Generic Motor*

Response). A between-subjects Welch's ANOVA revealed RT Cost differences (Welch's *F*(3, 52.7) = 6.59, *p* = .00073, est. $\omega^2 = 0.14$). Unlike Experiment 1, the RT Cost measure was somewhat sensitive to differences in audio-visual-motor alignment across groups. The baseline *Category Match: Location-Response (no color)* condition had a larger RT Cost than the *Category Mismatch: Color-Response* condition (*p* = .034), but there were no differences between *Category Match: Location-Response (no color)* versus *Category Match: Color-Response* (*p* = .87) or *Category Mismatch: Color-Response* versus *Category Match: Color-Response* (*p* = .11). Also unlike Experiment 1, the RT Cost was sensitive to motor demands: *Category Match: Location-Response (no color)* condition had a larger RT Cost than the *Generic Motor Response* condition (*p* = .029).

Generalization Test Accuracy

Generalization of incidental category learning was similar across Experiments 1 and 2 with the exception of the *Generic Motor Response* and *Passive Observation* conditions (Figure 4B). Notably, there was no evidence of generalization of learning in *Generic Motor Response* (*M* = 25.4%; *t*(25) = 0.21, *p* = .84, *d* = 0.040, 95% CI [21.0, 29.9]) or *Passive Observation* conditions (*M* = 31.1%; *t*(27) = 1.97, *p* = .059, *d* = 0.37, 95% CI [24.7, 37.5]).

Aligned with the results of Experiment 1, there was successful generalization of incidental learning in baseline *Category Match: Location-Response (no color)* (*M* = 51.0%; *t*(27) = 5.90, *p* < .001, *d* = 1.12, 95% CI [42.0, 60.0]) and *Category Match: Color-Response* (*M* = 35.7%; *t*(24) = 2.71, *p* = .012, *d* = 0.54, 95% CI [27.6, 43.9]). Generalization in the new Experiment 2 *Category Match: Location-Response* condition was also successful (*M* = 42.3%; *t*(24) = 3.97, *p* < .001, *d* = .79, 95% CI [33.3, 51.3]). Across conditions, when categories were *aligned* with visuomotor demands of the SMART task, there was successful generalization of incidental category learning.

By contrast, when auditory categories were *misaligned* with the task-relevant visual feature (color or location), there was no evidence of learning. This was true for both *Category Mismatch: Color-Response* (M = 23.7%; t(27) = -0.66, p = .51, d = 0.12, 95% CI [19.8, 27.7]) and *Category Mismatch: Location-Response* (M = 21.1%; t(27) = -2.43, p = 0.022, d = 0.46, 95% CI [17.7, 24.4]).

Generalization accuracy differed across the five replication conditions (Welch's F(4, 63.2) = 9.30, p < .001, est. ω^2 = 0.20). In line with observations from the RT Cost measure, the degree to which auditory-visual-motor information aligned impacted post-test accuracy. Most notably, and in contrast to Experiment 1, unique motor response demands had a strong effect on learning. The baseline *Category Match: Location-Response (no color)* condition had higher post-test accuracy than both the *Generic Motor Response* (p < .001) and *Passive Observation* conditions (p = .010), which did not differ significantly from one another (p = .74).

Post-test accuracy did not differ according to the visual feature that mapped to auditory categories and response (baseline *Category Match: Location-Response (no color)* versus *Category Match: Color-Response* conditions, p = .15). Again, the category-response mismatch mattered; baseline *Category Match: Location-Response (no color)* elicited more accurate generalization than *Category Mismatch: Color-Response* (p < .001). However, there was no difference between *Category Mismatch: Color-Response* versus *Category Match: Color-Response* conditions (p = .12).

The full crossing of visual features (location, color) with conditions in Experiment 2 allowed us to examine if incidental learning outcomes difference as a function of visual feature. They did not. There were no RT Cost differences between *Match* conditions (*Color-Response* vs. *Location-Response*, t(38.6) = 1.25, p = .22, d = 0.35, 95% CI [-19.3, 82.5]) or *Mismatch* conditions

(*Color-Response* or *Location-Response*, $t(45.9) = 0.43$, $p = .67$, $d = 0.11$, 95% CI [-9.24, 14.2]). Nor were there differences in generalization test accuracy (*Match*: $t(47.5) = -1.13$, $p = .27$, $d = 0.32$, 95% CI [-18.5, 5.21]; *Mismatch* $t(52.6) = 1.067$, $p = .29$, $d = 0.29$, 95% CI [-2.36, 7.72]).

In line with Experiment 1, in these new conditions, the alignment of auditory-visual-motor information impacted post-test accuracy (Welch's $F(2, 39.6) = 24.9$, $p < .001$, est. $\omega^2 = 0.38$). Specifically, post-hoc tests indicated poorer accuracy in the *Category Mismatch: Location-Response* condition versus either *Category Match: Location-Response (no color)* ($p < .001$) or *Category Match: Location-Response* ($p = .001$), which were not different from one another ($p = .85$). Alignment of auditory-visual-motor associations is a strong factor in driving incidental auditory category learning.

Together, these results support and extend conclusions from Experiment 1. When auditory categories incidentally align with unique motor responses to a primary visuomotor task that is ostensibly unrelated to those categories, learning is robust (Baseline, Category Match conditions), Without this relationship, the categories are not learned (Category Mismatch conditions). Notably, this is true even when there is a perfectly predictive association of the categories to a visual feature Overall, as illustrated in Figures 3 and 4, Experiment 2 replicated the in-laboratory results of Experiment 1 with online testing and randomized assignment to learning conditions, with an interesting exception. *Generic Motor Response* and *Passive Observation* elicited learning in Experiment 1, but not Experiment 2. Experiment 3 further examines this discrepancy.

Experiment 3

In Experiment 1, we observed successful learning in *General Motor Response* and *Passive Observation*. In Experiment 2, generalization performance did not differ from chance for either of these conditions. The online Experiment 2 replication was conducted in a very different setting than Experiment 1, with distinct samples. In Experiment 3, we sought to better understand the differences in learning outcomes with another replication experiment with random assignment of online participants to *General Motor Response* and *Passive Observation.*

Method

Participants

Participants were recruited via Prolific (www.prolific.sc) and participated using the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2019) in 1-day in May 2021. Participants were randomly assigned to one of the two conditions. Across conditions, there were 60 participants (36 M, 23 F, 1 non-binary) ages 18-35 years who were paid $10/hour for participating. There were 30 (17 M, 12 F, 1 non-binary) participants in *Generic Motor Response* and 30 (19 M, 11 F) in *Passive Observation*. Participants all passed a headphone check prior to beginning the experiment (Milne et al., 2020). Due to poor performance on catch trials (see above), a total of 10 participants were withheld from analyses (4 from the *Generic Motor Response* and 6 from *Passive Observation*). Included participants were 98.6% accurate on catch trials and excluded participants were 59.4% accurate on catch trials. All participants reported normal or corrected-to-normal vision and normal hearing.

Procedure

The stimuli and procedure were identical to *General Motor Response* and *Passive Observation* conditions of Experiment 2.

Results

Reaction Time Measures

Pre-processing. As in Experiment 2 *Generic Motor Response* trials for which there was a visual detection error or reaction time was less than 50 ms or greater than 1500 ms were excluded from analyses. This led to removal of 3.40% of trials. As in Experiments 1 and 2, including all trials did not change the outcomes of any analyses.

De-meaned Reaction Times. As in Experiments 1 and 2, we plotted the average RTs (Figure 3D) and the de-meaned RTs across blocks for the *Generic Motor Response* condition (Figure 3E).

Reaction Time Cost. According to a paired-samples *t*-test, and in line with trends from Experiments 1 and 2, participants in the *Generic Motor Response Condition* were significantly *faster* in Block 4 relative to Block 3 (*M* = -19.0 ms; *t*(25) = -3.60, *p* = .0014, *d* = 0.71, 95% CI [-29.8, -8.12]; Figure 3F), a RT facilitation rather than a cost. This is likely indicative of continued visuomotor task learning and not category learning because randomization of category to location did not disrupt visuomotor performance.

Generalization Test Accuracy

Replicating the findings in Experiment 2, there was no evidence of incidental category learning in either *Generic Motor Response* (*M* = 25.6%, *t*(25) = 0.28, *p* = .78, *d* = 0.055, 95% CI [21.2, 30.0]; Figure 4B) or *Passive Observation* conditions (*M* = 27.5%, *t*(23) = 0.99, *p* = .33, *d* = 0.20, 95% CI [22.2, 32.8]). Accuracies were not significantly different across the two conditions (*t*(45.9) = 0.56, *p* = .57, *d* = -0.16, 95% CI [-8.62, 4.79]). Thus, Experiment 3 replicates Experiment 2 and strengthens confidence in the conclusion that incidental learning is not robustly observed in the absence of a *unique* motor response or under conditions of passive observation of auditory-visual statistical regularities.

General Discussion

Category learning in natural environments 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. This contrasts with the kinds of overt category training typically studied in laboratory experiments. 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. Within an online sample that may be more representative of the typical adult population than Carnegie Mellon undergraduates, we observed a consistent lack of learning in tasks with limited or no motor demands. This study complements and extends earlier studies (Gabay et al., 2015; Lim & Holt, 2011; Liu & Holt, 2011; Wade & Holt, 2005) and demonstrates that incidental learning is driven by consistent alignment of task-relevant visual features with the variable acoustic exemplars defining an auditory category.

Incidental category learning occurred reliably when acoustically variable category exemplars consistently aligned with visuomotor demands of the primary task, but not when they were misaligned. The presence of an additional irrelevant visual feature that was uncorrelated with the primary task demands neither supported nor harmed incidental learning. By contrast, learning did *not* occur when auditory categories were aligned consistently with one visual feature, but the motor response in the primary task was aligned with another, category-unaligned visual feature. The importance of a category-specific motor response for learning was underlined by the results of Experiments 2 and 3: category learning did not reliably occur during passive observation or when participants made a generic motor response. This lack of learning is striking. In the absence of category-linked motor engagement, online participants did not appear to internalize the deterministic relationship between the visual feature and the auditory category that facilitated learning in the other conditions. However, the in-lab participants of Experiment 1 did appear to take advantage of these associative links – a point we return to below.

Overall, consistent mapping between a unique motor target and the variable acoustic exemplars defining each category greatly facilitates learning, even without overt feedback or category task-relevance. On this point, the *Mismatch* conditions are informative. For *Mismatch* conditions, auditory categories aligned perfectly with one of the two visual features (either location or color). This strong statistical regularity might be expected to support learning, even across passive exposure. For example, the addition of an aligned visual cue can help infants and adults segment statistically coherent units from continuous streams of speech (Cunillera et al., 2010; Thiessen, 2010) and the alignment of auditory and visual regularities is an important component in multimodal statistical learning (Mitchel & Weiss, 2011). Indeed, the very category-location alignment present in the *Category Mismatch: Color-Response* condition robustly supported incidental auditory category learning in baseline *Category Match: Location Response (No Color)*. Category learning was equally robust when the auditory categories were bound with location *or* color information – and, importantly, when there was *category-irrelevant* variation in a different visual feature, as shown in the *Match* conditions. Yet, category learning did not occur in the *Mismatch* conditions, where precisely the same visual grouping information was available. Why? Crucially, in the *Mismatch* conditions the category-relevant visual feature was decoupled from the motor task, which instead was linked to the category-uninformative visual feature.

A consistent mapping from an auditory category to a *unique* motor response also appears important for incidental learning. *Generic Motor Response*, for which participants responded to visual targets with a generic (spacebar) keypress, resulted in no auditory category learning across two replications (Experiments 2 and 3, conducted online with random assignment). This is notable, as this condition was identical to baseline *Category Match: Location Response (No Color)* that produced robust incidental category learning, but for the category-unique (four button) versus generic (spacebar) motor response to the visual target. *Generic Motor Response* had a lawful relationship between auditory categories and visual location, exactly as in *Category Match: Location-Response*. But participants failed to make use of the category-to-location relationship to learn the auditory categories when only a generic response was required. In other words, alignment of all categories to a single response discouraged learning – despite the availability of a consistent mapping from category to visual target location. The results of *Passive Observation* (Experiments 2 and 3) underscore these findings: when participants experienced the association between auditory categories and visual target locations passively, without a response, there was no incidental category learning.

The *Generic Motor Response* and *Passive Observation* results described above replicated across independent samples in Experiments 2 and 3, with no significant learning. Yet, the learning was observed in the same contexts in Experiment 1. We speculate that this rather puzzling discrepancy may be driven by participant sample and testing environment. Experiment 1 was conducted among Carnegie Mellon University undergraduates enrolled in psychology coursework, a population that others have regarded as distinct from the general population of learners (Henrich et al., 2010) – a WEIRD (Western, Educated, Industrialized, Rich and Democratic) sample. Adding to the weirdness, the Experiment 1 participants learned under conditions that amount to sensory deprivation – seated alone in a dimly lit and visually homogeneous sound-attenuated booth wearing noise-shielding headphones. This environment provides for excellent experimental control, but it presents a highly unusual, focused, context for learners to zero in on the only engaging input – our stimuli. It is possible that, in the context of a relatively undemanding *Passive Observation* and *Generic Motor Response* tasks, Experiment 1 participants – students in psychology courses – may have been more likely to engage in strategic hypothesizing about the study purpose than would the heterogeneous sample of online participants in Experiments 2 and 3 who participated at home. If this were the case, then it should give us pause about whether ‘passive’ observation evokes different cognitive demands across different participant samples and testing approaches. Savvy university student participants may be exploiting somewhat different strategies to discover patterns in input than experimenters expect under ‘passive’ or less-demanding conditions. Indeed, the online participants may reflect a broader sampling of the general population, which is a clear benefit of online recruitment methods that allow for sampling outside of local university populations (Clifford & Jerit, 2014). Beyond sampling, the conditions under which the online participants may have completed the experiment (i.e., at home with distractions rather than a sound-isolating booth) may more closely reflect real-world learning conditions, such as during everyday speech perception. Nonetheless, given the robust replication with online participants and random assignment across Experiments 2 and 3, it seems safe to conclude that incidental category learning is more fragile – and maybe absent – under *Passive Observation* or *Generic Motor Response* conditions that destroy the unique category-to-response mapping.

A unique mapping from category to response, then, seems to be the major factor in driving incidental auditory category learning. Yet, one might wonder if category-unique motor responses are beneficial due to this unique motor mapping – or simply because this more differentiated response may draw more engagement and attention to the task than just pressing a spacebar when seeing an X (*Generic Motor Response*), or just observing those Xs appear on the screen (*Passive Observation*). However, on several grounds, this is not a particularly compelling explanation for the lack of learning we observed online in these conditions. First, participants' catch trial performance was excellent (97% average accuracy for all included participants), indicating attentiveness. Second, the spacebar reaction times for the *Generic Motor Response* were not only very fast, but also showed particularly low variability – the opposite of what we would expect from inattentive or distracted participants (Kofler et al., 2013; Gómez-Guerrero et al., 2011).

Nonetheless, the mapping to motor responses may not be as critical for learning some types of categories, under some conditions (Ashby et al., 2003). For example, it is possible to learn categories associated with a unique visual cue that is independent of the response mapping when participants are aware of the categories and use feedback to learn them explicitly. For instance, one category of images is mapped to a red circle and another category is mapped to a blue circle, people can learn even when the categories are associated with a different response key on each trial (Spiering & Ashby, 2008). This suggests that in overt learning contexts – where it is clear to participants that they are performing a category-learning task – learners can overcome inconsistent category-response mappings when a consistent visual cue is present. However, such overt tasks are not encountered frequently: typically, people are quite unaware of how auditory, visual, and motor information might relate to one another, much less to regularities defining novel categories.

The cross-condition differences in mean RT also bear examination with regard to interpreting the role for motor responses. Although mean RT did not influence RT Cost or post-test accuracy measures, it is obvious that visual feature and genre of motor response had a marked impact on mean RT. Across conditions, responses were fastest in *Generic Motor Response conditions,* which simply required a tap of the spacebar to any X that appeared. Responses were slowest for conditions mapping motor response to color and intermediate for mapping motor response to one of four locations. The longer RTs for responses to color were likely due to the arbitrary and newly learned mapping of motor response to color. By contrast, spatial location intuitively maps to the left-to-right keyboard assignment. Nonetheless, these mean RT differences were *not* accompanied by commensurate differences in categorization accuracy.

The representational glue for incidental category learning

The temporal alignment of auditory, visual, and motor responses has been argued to act as ‘representational glue’ that binds acoustically variable category exemplars – doing so without instructions, overt category decisions or feedback (Gabay et al., 2015; Lim & Holt, 2011; Wade & Holt, 2005). The present results refine this perspective. Here, experiencing a (perfect) statistical regularity between auditory category and visual feature was not sufficient for robust learning without alignment with a category-unique motor response.

At this juncture, it is helpful to consider incidental learning in broader terms. Incidental category learning as instantiated in the present study may serve as a useful experimental proxy for 'real-world' settings where learners are active and can capitalize on rich associations that exist between category exemplars and other objects, events, and their own behaviors. In this context, we highlight five points. First, although there is abundant evidence that organisms – both human and nonhuman – are able to learn statistical regularities across mere exposure to the input (Frost et al., 2019; Saffran & Kirkham, 2018; Saffran et al., 1996), there are learning challenges for which passive exposure to regularities is insufficient to drive learning (Cristia, 2018; Emberson et al., 2013; Wade & Holt, 2005). The present results show that incidental learning hastens category acquisition above and beyond passive exposure, without the need for overt category decisions or explicit feedback. Specifically, Experiment 2 (replicated by Experiment 3) showed no significant auditory category learning in *Passive Observation*, despite a deterministic relationship between visual location and acoustic category. It is possible that longer exposure might ultimately result in learning under *Passive Observation*. But, crucially, the present results demonstrate that simply embedding these same regularities in an active task unrelated to category learning produces robust learning across the same number of trials – as long as the active behavior in a task ostensibly unrelated to category learning tacitly aligns with category membership. Thus, incidental learning may hasten learning that is difficult or impossible through mere exposure alone.

The second issue involves how learners know *what* to learn. Statistical learning accounts have long grappled with the question of how the system selects which of the countless potential environmental regularities are to be learned. Incidental learning suggests that active behavior may serve to ‘funnel’ information. As we described above, passive accumulation of auditory statistical regularities is facilitated by alignment with visual events (Cunillera et al., 2010; Mitchel & Weiss, 2011; Thiessen, 2010). In the present study, the underlying acoustic regularities defining auditory categories aligned with visual location or color. Whether learners learned from the acoustic regularities to acquire auditory categories depended upon whether those regularities aligned with active behavior. The *Match* and *Mismatch* conditions possessed the same auditory-visual statistical regularities; only the alignment of active behavior with the regularities differed across conditions. Category learning was observed when behavior was directed at a task that aligned with the regularity. In this way, active engagement in an environment may encourage ‘foraging’ for information that directs learners to specific statistical regularities among the essentially infinite informational contingencies that exist in even simple real-world environments.

The third issue relates to the ‘representational glue’ that binds variable input exemplars together as categories. One might have expected *Generic Motor Response* learning to exceed *Passive Observation* learning because the temporally aligned motor response for all trials would encourage trial-by-trial engagement to the information-bearing signal. Yet, Experiment 2 (replicated in Experiment 3) found no evidence of *Generic Motor Response* learning. Just as the category-unique motor response was effective at driving the learning system to form separate categories based on the auditory-visual-motor regularities, the single motor response may have been effective at driving the learning system to forage for the regularities aligned with the active one-button behavior. In this case, the ‘representational glue’ would bind *all* exemplars across all categories according to their overarching regularities. The *Generic Motor Response* might appear deleterious to learning precisely because it encourages incidental learning that collapses the experimenter-intended categories into a single representation. If this were the case, then one would expect poor performance on the four-alternative labeling task, as we observed. However, this would not be indicative of *no* incidental learning, it would be indicative of learning a single category representation that could not drive performance in the four-alternative task. Future studies that examine the perceptual space before and after incidental auditory category learning would be able to address this possibility.

The fourth issue relates to how within-category exemplar variability aligns with task-relevant factors in the primary task. In Gabay et al. (2015), Block 4 trials were randomized such that each of the five auditory stimuli could be drawn from *any* of the four auditory categories (*category-mixed* trials). Therefore, not only did the sounds convey no information about target location in Block 4, there was also no consistent category membership across the five acoustic exemplars preceding a visual target in Block 4. The *Category Match: Location-Response (no color)* condition in Experiments 1 and 2 was a near replication of Gabay et al. (2015), except in execution of how Block 4 was randomized. In contrast to Gabay et al., exemplars from the same auditory category defined a Block 4 trial, thus maintaining category exemplar similarity and coherence within a trial even as the category-to-location mapping was randomized. Examining outcomes across studies, these differences appear to impact learning. Whereas Gabay et al. report a 77 ms RT Cost, the similar baseline condition in Experiment 1 (20.8 ms) and Experiment 2 (27.4 ms) had substantially smaller RT Costs. Whereas Gabay et al. report average post-test accuracy of

65.8%, the baseline condition in Experiment 1 (47.6%) and Experiment 2 (51.0%) had lower accuracies. Although further investigation is warranted in advance of strong conclusions (especially in the context of possible cohort differences), this divergence might suggest that learners are sensitive to both the category-to-location-to-response mapping in the SMART task that produces the RT Cost and also to within-category regularities across exemplars. Eliminating the within-category regularity, as in the Gabay et al. manipulation, came at a much greater cost to the efficiency of visual target detection.

Finally, it is worth considering variability of a different sort – that of participants' learning outcomes. Learning across perceptual input invariably produces individual differences in learning outcomes (Roark & Chandrasekaran, 2021; Shamloo & Hélie, 2020; Shen & Palmeri, 2016). In the absence of sounds during the SMART task, participants get faster at detecting the visual targets with practice across blocks (Gabay et al., 2019). Thus, it is important to recognize that incidental auditory category learning travels together with visuomotor learning of task demands. This insight might be helpful in considering participant variability in RT Cost. Participants may vary in the extent to which they are influenced by the sound categories and their incidental relationship to the primary visuomotor task. Those participants who are more influenced would be considered 'good learners' in the present study inasmuch as they learned our target of interest: auditory categories. This is evidenced by positive RT Costs (slowing of visuomotor behavior) upon disruption of the category-to-task relationships, due to reliance on newly acquired categories in driving behavior. Other participants would here be considered 'poor learners' in this context as they show no RT Cost, or even a speeding of RT in Block 4. The heightened variability in RT Costs in the *Category Match: Color-Response* condition across both Experiments 1 and 2 may be driven by how well participants learned both the auditory categories and the color-response mapping, which was a more demanding visuomotor task than the location-response mapping. Individuals may differ in the extent to which they 'cast a wide net' in foraging for new regularities with some participants more susceptible to incidental alignment of audio-visual regularities than others.

Conclusion

Prior research has focused on category learning via overt feedback (Ashby & Maddox, 2011; Francis & Nusbaum, 2002; Goldstone, 1994; McCandliss et al., 2002; McClelland et al., 2002; Nosofsky, 1986) or learning across passive exposure or without supervision (Ashby et al., 1999; Clapper & Bower, 2002; Ell et al., 2012; Folstein et al., 2010; Goudbeek et al., 2009; Kaplan & Murphy, 1999; Maye et al., 2002; McMurray et al., 2009; Vallabha et al., 2007). In contrast, in incidental learning categories are discovered via their utility in supporting behavior on a primary task. Incidental tasks model learning 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 – and yet, are not entirely passive observers. The present results demonstrate that incidental category learning is supported by alignment between the variable acoustic exemplars defining a category and a unique behavioral response, even when that response is directed at a task ostensibly unrelated to the categories. In this way, active behavior provides the representational glue that supports incidental acquisition of categories across statistical regularities in the input.

References

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our Midst: An online behavioral experiment builder. *Behavior Research Methods*, 438242. https://doi.org/10.3758/s13428-019-01237-x

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

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*(7), 1114–1125. https://doi.org/10.3758/bf03196132

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*(6), 1178–1199. https://doi.org/10.3758/bf03207622

Clapper, J. P. (2012). The effects of prior knowledge on incidental category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(6), 1558–1577. https://doi.org/10.1037/a0028457

Clapper, J. P., & Bower, G. H. (2002). Adaptive categorization in unsupervised learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 908–923. https://doi.org/10.1037/0278-7393.28.5.908

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Clifford, S., & Jerit, J. (2014). Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies. *Journal of Experimental Political Science*, *1*(2), 120. https://doi.org/10.1017/xps.2014.5

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, *170*(2002), 312–327. https://doi.org/10.1016/j.cognition.2017.09.016

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology*, *63*(2), 260–274. https://doi.org/10.1080/17470210902888809

Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *The Quarterly Journal of Experimental Psychology*, *65*(8), 1537–1562. https://doi.org/10.1080/17470218.2012.658821

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, *128*(1), 82–102. https://doi.org/10.1016/j.cognition.2012.12.006

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

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. https://doi.org/10.1037//0096-1523.28.2.349

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical Learning Research: A Critical Review and Possible New Directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Gabay, Y., Dick, F. K., Zevin, J., & Holt, L. L. (2015). Incidental Auditory Category Learning. *Journal of Experimental Psychology: Human Perception and Performance Learning*, *41*, 1124–1138. https://doi.org/10.1037/xhp0000073

Gabay, Y., Karni, A., & Holt, L. L. (2019). Overnight consolidation and retention of implicit and explicit knowledge of incidentally learned auditory categories. Interdisciplinary Advances in Statistical Learning, San Sebastian, Spain.

Goldstone, R. L. (1994). Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.

Gómez-Guerrero, L., Martín, C. D., Mairena, M. A., Martino, A. D., Wang, J., Mendelsohn, A. L., Dreyer, B. P., Isquith, P. K., Gioia, G., Petkova, E., & Castellanos, F. X. (2011). Response-Time Variability Is Related to Parent Ratings of Inattention, Hyperactivity, and Executive Function. *Journal of Attention Disorders*, *15*(7), 572–582. https://doi.org/10.1177/1087054709356379

Goudbeek, M., Swingley, D., & Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1913–1933. https://doi.org/10.1037/a0015781

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Holt, L. L. (2011). How perceptual and cognitive constraints affect learning of speech categories. In A. Cohn, C. Fourgeron, & M. Huffman (Eds.), *Handbook of Laboratory Phonology*. Oxford University Press.

Idemaru, K., & Holt, L. L. (2011). Word Recognition Reflects Dimension-based Statistical Learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956. https://doi.org/10.1037/a0025641

Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, *27*(4), 699–712. https://doi.org/10.3758/bf03211563

Kimball, G., Cano, R., Feng, J., Feng, L., Hampson, E., Li, E., Christel, M. G., Holt, L. L., Lim, S., Liu, R., & Lehet, M. (2013). Supporting Research into Sound and Speech Learning through a Configurable Computer Game. *2013 IEEE International Games Innovation Conference (IGIC)*, 110–113. https://doi.org/10.1109/igic.2013.6659172

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, *33*(6), 795–811. https://doi.org/10.1016/j.cpr.2013.06.001

Lim, S.-J., Fiez, J. A., & Holt, L. L. (2019). Role of the striatum in incidental learning of sound categories. *Proceedings of the National Academy of Sciences*, *116*(10), 201811992. https://doi.org/10.1073/pnas.1811992116

Lim, S.-J., & Holt, L. L. (2011). Learning Foreign Sounds in an Alien World: Videogame Training Improves Non-Native Speech Categorization. *Cognitive Science*, *35*(7), 1390–1405. https://doi.org/10.1111/j.1551-6709.2011.01192.x

Lim, S.-J., Lacerda, F., & Holt, L. L. (2015). Discovering functional units in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1139–1152. https://doi.org/10.1037/xhp0000067

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*(3), 683–698. https://doi.org/10.1162/jocn.2009.21392

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review*, *9*(4), 829–835. https://doi.org/10.3758/bf03196342

Markman, A. B., & Ross, B. H. (2003). Category Use and Category Learning. *Psychological Bulletin*, *129*(4), 592–613. https://doi.org/10.1037/0033-2909.129.4.592

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111. https://doi.org/10.1016/s0010-0277(01)00157-3

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience*, *2*(2), 89–108. https://doi.org/10.3758/cabn.2.2.89

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiology & Behavior*, *77*, 657–662. file:///Users/devans/Documents/Papers2/Articles/2003/Unknown/2003 R8705.pdf%5Cnpapers2://publication/uuid/D9D9D273-E580-4543-BB39-F6DA81E6B21F

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 1–12. https://doi.org/10.3758/s13428-020-01514-0

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1081–1091. https://doi.org/10.1037/a0023700

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision*, *16*(3), 3–3. https://doi.org/10.1167/16.3.3

Protopapas, A., Mitsi, A., Koustoumbardis, M., Tsitsopoulou, S. M., Leventi, M., & Seitz, A. R. (2017). Incidental orthographic learning during a color detection task. *Cognition*, *166*, 251–271. https://doi.org/10.1016/j.cognition.2017.05.030

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 75–94. https://doi.org/10.1002/wcs.1268

Roark, C. L. & Chandrasekaran, B. (2021). Individual variability in strategies and learning outcomes in auditory category learning. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar (Eds.), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society,* Austin, TX: Cognitive Science Society.

Roark, C. L., & Holt, L. L. (2018). Task and distribution sampling affect auditory category learning. *Attention, Perception, & Psychophysics*, *80*(7), 1804–1822. https://doi.org/10.3758/s13414-018-1552-5

Roark, C. L., & Holt, L. L. (2019). Perceptual dimensions influence auditory category learning. *Attention, Perception, and Psychophysics*, *81*(4), 912–926. https://doi.org/10.3758/s13414-019-01688-6

Roark, C. L., Lehet, M., Dick, F., & Holt, L. L. (2020, October 28). The representational glue for incidental category learning is alignment with task-relevant behavior. https://doi.org/10.17605/OSF.IO/9DKJG

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., & Kirkham, N. Z. (2018). annurev-psych-122216-011805.pdf. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Scharinger, M., Henry, M. J., & Obleser, J. (2013). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Memory & Cognition*, *41*(5), 752–768. https://doi.org/10.3758/s13421-013-0294-9

Seitz, A. R., Protopapas, A., Tsushima, Y., Vlahou, E. L., Gori, S., Grossberg, S., & Watanabe, T. (2010). Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. *Cognition*, *115*(3), 435–443. https://doi.org/10.1016/j.cognition.2010.03.004.unattended

Shamloo, F., & Hélie, S. (2020). A study of individual differences in categorization with redundancy. *Journal of Mathematical Psychology*, *99*, 102467. https://doi.org/10.1016/j.jmp.2020.102467

Shen, J., & Palmeri, T. J. (2016). Modelling individual difference in visual categorization. *Visual Cognition*, *24*(3), 260–283. https://doi.org/10.1080/13506285.2016.1236053

Spiering, B. J., & Ashby, F. G. (2008). Response processes in information–integration category learning. *Neurobiology of Learning and Memory*, *90*(2), 330–338. https://doi.org/10.1016/j.nlm.2008.04.015

Thiessen, E. D. (2010). Effects of Visual Information on Adults' and Infants' Auditory Statistical Learning. *Cognitive Science*, *34*(6), 1093–1106. https://doi.org/10.1111/j.1551-6709.2010.01118.x

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General*, *141*(2), 363–381. https://doi.org/10.1037/a0025014

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*(4), 2618–2633. https://doi.org/10.1121/1.2011156

Wiener, S., Murphy, T. K., Goel, A., Christel, M. G., & Holt, L. L. (2019). Incidental learning of non-speech auditory analogs scaffolds second language learners' perception and production of Mandarin lexical tones. *Proceedings of the International Congress of Phonetic Sciences*.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional Phonetic Learning at 10 Months of Age. *Infancy*, *15*(4), 420–433. https://doi.org/10.1111/j.1532-7078.2009.00024.x