Proceedings of the Annual Meeting of the Cognitive Science Society

Perceptual Category Learning Results in Modality-Specific Representations

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Roark, Casey L

2024

Perceptual Category Learning Results in Modality-Specific Representations

Casey L. Roark (casey.roark@unh.edu)
Department of Psychology, University of New Hampshire
15 Academic Way, Durham, NH 03824 USA

Abstract

Categorization is a fundamental cognitive skill that spans the senses. Even so, most research has focused on categorization and category learning in the visual modality. As a result, it is not yet clear how modality influences the perceptual and cognitive processes supporting category learning. In two experiments, we tested whether category learning results in amodal or modality-specific representations. We found strong evidence for modality-specific representations with independent learning across modalities. These results highlight the need to look beyond vision when constructing and testing models categorization and category learning. These findings also contribute to the longstanding debate on the amodal/modal nature of human knowledge, which is of broad interest to the cognitive science community.

Keywords: category learning; audition; vision; modality; domain-general; abstraction

Introduction

Auditory and visual information are fundamental to typical perception of the world. Perceptual categories built on auditory and visual information can be useful ways to organize the world. For example, based on visual features, one might rapidly identify an approaching creature as friend or foe (e.g., dog vs. wolf). Based on auditory features, we can rapidly identify the sounds and words in speech to effectively communicate (e.g., *b*oysenberry vs. *p*oison berry).

Despite the ubiquity of categories both across and within different sensory modalities, much of what we know about perceptual categorization and category learning comes from the visual modality. As a result, it is not yet clear how modality influences the perceptual and cognitive processes supporting category learning. Of particular interest to the broader cognitive science community is whether the representations that are formed during learning are abstract and amodal or whether they are specific to the modality and even stimuli being learned (Kaup et al., 2023; Machery, 2016; Reed, 2016). Here, we test whether representations acquired during artificial auditory and visual category learning are amodal or modality specific.

One clear finding in categorization research is that humans can relatively quickly learn categories based on rules that require selective attention to individual dimensions or features (Shepard, Hovland & Jenkins, 1961). For example, one could arrange the objects in Figure 1 on the basis of shape (e.g., circles belong to category A, triangles belong to category B) or color (e.g., black shapes belong to category A, white shapes belong to category B). Regarding these so-called ‘rule-based’ categories, learners can form such abstract rules to learn categories. These results were initially demonstrated and examined almost exclusively in vision (Ashby & Maddox, 2011) and have been recently examined audition (Chandrasekaran et al., 2013; Roark et al., 2021).

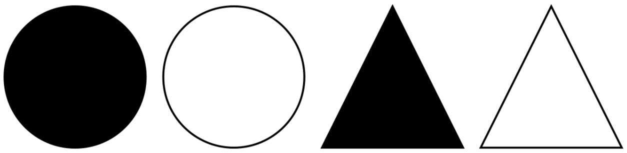

Figure 1: Object examples.

However, indirect comparisons of auditory and visual learning have highlighted important ways in which learning differs across modalities. For example, while many visual dimensions are separable and therefore reasonably easy to create these rules that require selective attention, many auditory dimensions are integral and difficult to create such rules (Garner, 1974). As a result, while learning visual categories that can be easily described by rules is typically very fast in adult humans (Ashby & Maddox, 2011; Shepard et al., 1961), this result is true for some auditory categories (Goudbeek et al., 2009; Scharinger et al., 2013), but not all (Roark & Holt, 2019; Roark, Plaut, & Holt, 2022).

There are very few direct comparisons of auditory and visual category learning. However, several recent studies (Roark et al., 2021, 2023; Roark, Smayda, & Chandrasekaran, 2022) have directly compared auditory and visual learning using psychologically (Visscher et al., 2005) and neurally (Schönwiesner & Zatorre, 2005) comparable stimulus dimensions. These studies have found that even when overall learning performance (i.e., accuracy) is the same, learners may employ distinct decision processes during auditory and visual category learning (Roark et al., 2021), that learning is affected by modality-specific expertise (i.e., musicianship; Roark, Smayda, et al., 2022), and that the advantage of learning in adults over children is much more profound in problems that seem to mirror naturalistic learning contexts that are specific to modality (e.g., rules in vision, similarity in audition; Roark et al., 2023). Therefore, it is of significant theoretical interest to understand whether perceptual category learning is supported by shared, amodal mechanisms or independent, modality-specific mechanisms.

The amodal or modality-specific nature of knowledge representation (such as that acquired during category learning) is a fundamental puzzle in cognitive science, with many different perspectives across philosophy, psychology, and neuroscience (Barsalou et al., 2003; Bimbard et al., 2022; Binder, 2016; Ghazanfar & Schoreder, 2006; Kaup et al., 2023; Kiefer & Pulvermüller, 2012; Machery, 2016; Mahon & Caramazza, 2008; Reed, 2016). This matter is certainly not

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

yet resolved. However, the question of whether learned representations are amodal or modality-specific has been addressed in another area of learning – implicit statistical learning (Conway & Christiansen, 2005, 2006; Frost et al., 2015; Mitchel & Weiss, 2011; Seitz et al., 2007). In these studies, learners simultaneously learned regularities from auditory and visual information streams. Rather than forming a shared, amodal representation of the regularities they encountered across modalities, learners created modality-specific representations and simultaneously tracked auditory and visual signals as separate sources of information. These results indicate that modality is likely retained as a feature in our internal representation of our external sensory world.

Prior research on implicit statistical learning makes a clear argument for the modality-specificity of at least some aspects of learning. However, it is still unclear whether other forms of learning are similarly affected by such modality-specificity. In the current study, we examine the nature of representations formed during overt, supervised learning of visual and auditory categories. The main goal of the current study is to determine whether simultaneous interleaved auditory and visual category learning leads to amodal or modality-specific representations (vs. blocked). Specifically, we aimed to understand whether learners could exploit a shared higher-level regularity – a rule – across modalities to build abstract, amodal representations.

Amodal versus modality-specific frameworks

In the current study, we will leverage a simple experimental manipulation (blocked versus interleaved trials across modalities) to understand the nature of the representations formed during perceptual category learning. When auditory and visual categories are learned in completely separate blocks, learners may default to constructing modality-specific representations and treat the category tasks as completely separate. In contrast, when auditory and visual categories are learned simultaneously within the same task (i.e., interleaved), we can assess conflicting predictions about whether learners will construct one abstract, amodal representation across modalities (e.g., "Categories can be separated by a rule on one stimulus dimension – category A exemplars are high that dimension, category B exemplars are low on that dimension") or two grounded, modality-specific representations in separate modalities (e.g., "Auditory categories can be separated by a rule on temporal modulation – category A exemplars have high temporal rates, category B exemplars have low temporal rates" and "Visual categories can be separated by a rule on spatial frequency – category A exemplars have high spatial frequencies, category B exemplars have low spatial frequencies"). An amodal representation thus would abstract away modality-specific information to acknowledge the shared rule-based structure across modalities, while a modality-specific representation would retain information only within the specific modality.

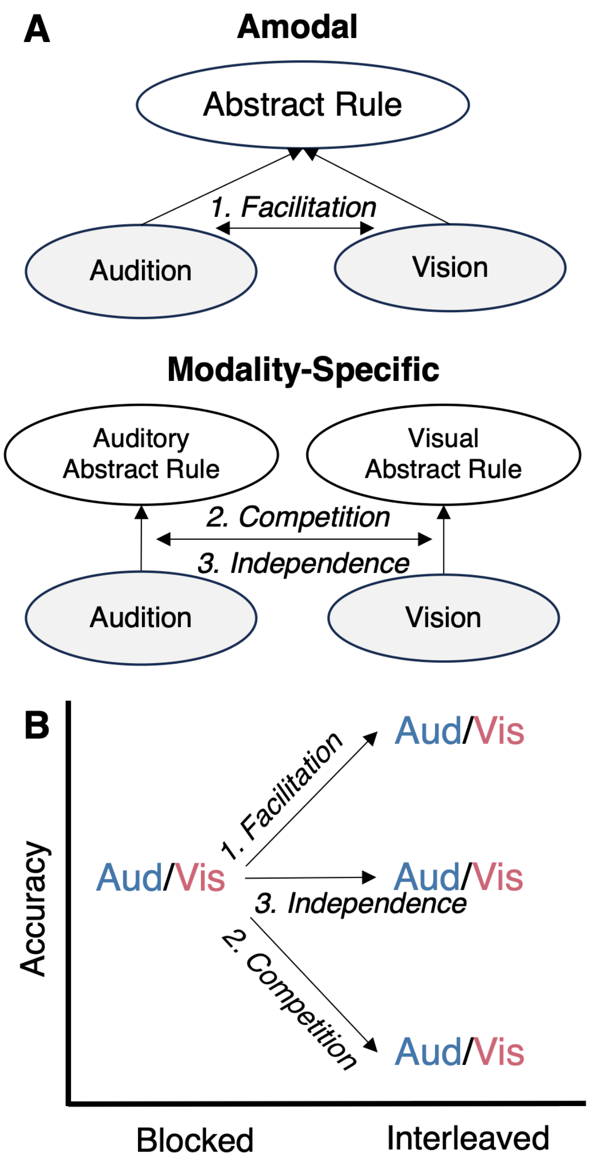

Figure 2: **A.** Theoretical frameworks **B.** Learning performance predictions according to the frameworks.

Figure 2A illustrates the competing amodal and modality-specific theoretical frameworks. According to an **amodal representation** framework, if auditory and visual categories share the same high-level regularities and can be learned with the same domain-general rule (e.g., "Categories can be separated by a rule on one stimulus dimension – category A exemplars are high that dimension, category B exemplars are low on that dimension"; Figure 3), then a shared, amodal abstract category representation should be formed. If simultaneous auditory and visual category learning supports an amodal representation, there should be a *facilitation effect,* such that simultaneous learning with interleaved trials of different sensory modalities will lead to enhanced performance relative to separate blocked training as learners can use what they have learned in one modality to facilitate their learning in the other modality (Figure 2B).

In contrast, according to a **modality-specific representation** framework, even if auditory and visual categories share the same high-level regularities and rule structure, learning will still be impacted by modality-specific constraints. There are two possible effects that would be consistent with a modality-specific framework. First, there may be a *competition effect* between modalities, such that interleaved learning in two modalities will strain general working memory and attentional resources as participants need to switch between two seemingly unrelated tasks, limiting learning in both modalities. If there is competition between modalities, then performance in both sensory modalities should be impaired in interleaved relative to blocked training (Figure 2B). This prediction is generally in line with viewing modality-interleaved training as a kind of task switching paradigm in which participants must switch their attention between competing tasks across trials (e.g., Crossley et al., 2023; Kiesel et al., 2010; Monsell, 2003). Such task switching generally decreases overall learning, when investigated within modality (Crossley et al., 2023).

Second, a modality-specific perspective may also predict an *independence effect* between modalities, such that learning categories from two modalities at once may occur completely independently and thus interleaved will not be different from blocked training (Figure 2B). If this is the case, it would indicate that learners may have separate pools of resources and separate rules that they may draw upon to simultaneously learn the auditory and visual categories. This prediction is in line with the findings from the implicit statistical learning literature that find similar performance when auditory and visual sequence regularities are presented simultaneously or separately (Christiansen & Conway, 2005, 2006).

Exp 1: Blocked vs. interleaved learning

Experiment 1 compares blocked to interleaved auditory and visual category learning to test whether interleaved learning across modalities supports shared amodal representations leading to facilitation of learning in the two modalities *or* whether interleaved learning supports modality-specific representations leading to either competition or independence of learning in the two modalities.

Method

Participants Participants were 119 undergraduate students from the University of New Hampshire, ages 18-23 (102 F, 17 M), who participated for partial course credit. Participants completed all tasks in person using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants learned two auditory and two visual categories with the same category structure that could be learned with the same domain-general rule (Figure 3). Participants were randomly assigned to learn categories across modalities either simultaneously (interleaved) or separately (blocked). All participants learned both auditory and visual categories. One participant was excluded due to experimenter error.

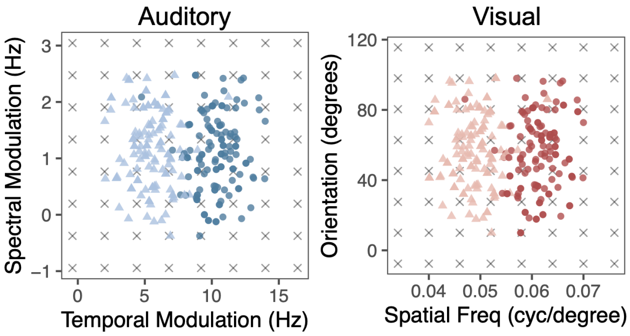

Figure 3: Category distributions.

Stimuli Participants learned auditory and visual categories comprised of fundamental auditory or visual dimensions that are psychological and neurally comparable (Schönwiesner & Zatorre, 2009; Visscher et al., 2007) and important for natural perception (i.e., speech perception, object recognition; Woolley et al., 2005). In the auditory modality, participants learned categories that were distinguished by the temporal modulation rate of a nonspeech sound, while sounds varied irrelevantly in spectral modulation. In the visual modality, participants learned categories that were distinguished by the spatial frequency of a Gabor patch, while images varied irrelevantly in orientation. Participants were not told about the dimensions that distinguish the stimuli.

Each category type had two individual categories. These specific stimuli and categories have been used before in studies of category learning (Roark et al., 2021, 2023) demonstrating that participants are able to learn these categories by creating rules along the category-relevant dimensions. Category distributions were initially generated in normalized space and then transformed separately into auditory and visual spaces. As such, the auditory and visual categories have the same distributions. Previous research has demonstrated that rule-based learning along temporal modulation and spatial frequency is comparable when learned in separate tasks (Roark et al., 2021). As such, we expected that blocked learning performance would be comparable across sensory modalities. Participants also completed a generalization test that was comprised of novel stimuli not encountered during training and formed a uniform grid across the stimulus spaces (Figure 3 – Xs). We explain the purpose of the generalization test below.

Procedure All participants gave informed consent, and all study procedures were approved by the Institutional Review Board at the University of New Hampshire. After consent procedures, participants completed a brief demographics questionnaire. To hear auditory stimuli, participants wore Seinheisser over-ear headphones. Participants were informed that they would be learning auditory and visual categories. On each trial, participants heard a sound or saw an image of a category exemplar randomly selected from the distribution. They responded which category they thought the stimulus belonged to using keyboard. Participants were given distinct labels and keypress responses for categories in different modalities to ensure minimal overlap (auditory: “1”, “2”; visual: “3”, “4”), which has been shown to be beneficial for interleaved learning within the same modality (Crossley et al., 2023). Participants were not told that the categories followed any rules or that there was any relationship between modalities.

In the blocked condition, participants completed 200 trials in 4 blocks of 50 trials each in one modality and then 200 trials in 4 blocks of 50 trials each in the other modality. Modality order was counterbalanced across participants. In the interleaved condition, participants completed 400 trials (200 in each modality) in 8 blocks of 50 trials each with short breaks in between. Auditory and visual trials were randomly interleaved.

After training, participants in both conditions completed a generalization test. In the generalization test, the auditory and visual modalities were presented in separate blocks (64 trials per block), with order counterbalanced across modalities. Participants were shown novel exemplars from the relevant categories and were told to respond based on the category labels they learned during training.

The generalization test was included to ensure that potential differences in performance between interleaved and blocked conditions was not due only to surface-level differences in the conditions (e.g., different buttons, higher working memory constraints in switching between tasks) and instead were due to differences in the amodal versus modality-specific representations. Because learners in the interleaved condition simultaneously needed to hold on to both auditory and visual information in both training and test, to create a fair comparison with blocked training, participants completed the generalization test blocks after both auditory and visual training were complete. The generalization test was always blocked by modality with order counterbalanced across participants. As such, participants in both blocked and interleaved conditions needed to hold on to both auditory and visual category representations to perform well in the generalization test. Participants in both conditions were informed at the beginning of the experiment that they would be tested on both auditory and visual categories at the end of the experiment.

Results

In line with prior work (Roark et al., 2021), auditory and visual performance was not significantly different when learned in separate blocks. This is true for both training ($t(58) = -1.26$, $p = .21$, $d = -0.16$) and generalization test (which was completed after training; $t(58) = -1.71$, $p = .093$, $d = -0.22$). When learning categories separately, accuracies were no different across modalities, indicating that the categories were well-matched for difficulty.

We next compared the relative performance of visual and auditory categories in the blocked and interleaved conditions (Figure 4). As a reminder, based on the theoretical perspectives we made three distinct predictions for interleaved performance relative to blocked based on the relationship between modalities – facilitation (amodal), competition (modality-specific), or independence (modality-specific). Our predictions are not different for training and test performance, but we examine both measures to ensure a fair comparison between blocked and interleaved conditions.

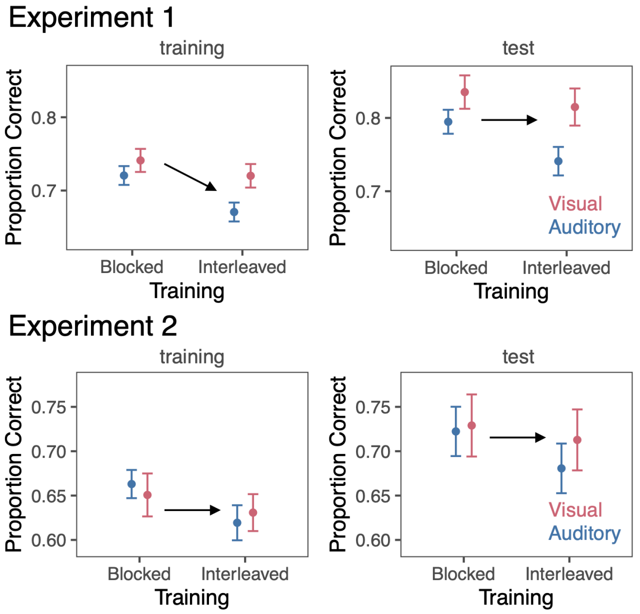

Figure 4: Learning performance in both experiments. Error bars reflect *SEM.* The black arrows represent the nature of the relationship between blocked and interleaved training based on the three competing predictions (facilitation: blocked < interleaved; competition: blocked > interleaved; independence: blocked n.s. interleaved).

During training, performance was significantly better in the blocked condition relative to the interleaved condition, supporting the competition prediction ($F(1, 116) = 4.51$, $p = .036$, $\eta_G^2 = 0.025$). However, in the generalization test which was blocked by modality for both conditions, there were no significant differences between blocked and interleaved conditions ($F(1, 116) = 2.42$, $p = .12$, $\eta_G^2 = 0.013$), supporting the independence prediction.

When examining blocked training alone, there was no significant differences between auditory and visual learning. However, when blocked training was combined with interleaved training, performance was significantly better in the visual than the auditory modality in both training ($F(1, 116) = 8.74$, $p = .004$, $\eta_G^2 = 0.025$) and test ($F(1, 116) = 9.78$, $p = .002$, $\eta_G^2 = 0.030$). In both training ($F(1, 116) = 1.48$, $p = .23$, $\eta_G^2 = 0.004$) and test ($F(1, 116) = 0.83$, $p = .36$, $\eta_G^2 = 0.003$), there were no significant interactions of training type and modality – indicating that the differences between modalities and training types did not depend on one another.

Regarding our predictions, these results support a modality-specific framework with competition effects in training (blocked > interleaved) and independent effects in test (blocked was not different from interleaved).

It is also possible that participants may have constructed a shared amodal representation of the categories even when learning was blocked by modality. Participants were not told about potential relationships between the rules across modalities and we reasoned that any relationship would be more evident to participants in the interleaved condition. However, it is still possible that participants did create some sort of amodal representation in the blocked condition. If learners were able to use the knowledge that they acquired in the first modality to facilitate learning in the second modality, we would expect performance to be better for the second task. That is, if learners could transfer their modality-specific knowledge to another modality, which may be supported by a shared amodal representation, we would expect performance to be better in the second compared to the first task. In contrast to this prediction, there were no significant differences in test performance when tasks were completed first and second (auditory ($t(52) = 1.30$, $p = .20$, $d = 0.34$; visual: $t(46) = -1.62$, $p = .11$, $d = -0.42$). These results indicate that in the blocked condition, knowledge acquired about the categories in one modality neither facilitated nor competed with knowledge acquired about the categories in the other modality. This also supports a modality-specific independence of representations – learners formed and applied category knowledge separately across modalities.

Overall, these results support a modality-specific representation framework of category learning wherein learners build separate, modality-specific representations even when learning auditory and visual categories simultaneously. It is clear that learners did *not* exploit the shared higher-level regularities using amodal representations to facilitate learning across modalities. However, the results are mixed with regard to competition and independence effects across modalities. One potential reason for these mixed results is the increased cognitive demand required to switch category labels and response keys across auditory and visual trials during interleaved training. It is possible that the competition effects observed during training may have been due to this greater demand. Additionally, it is possible that because auditory and visual categories relied on distinct category labels/response keys, all learners may have been discouraged from creating an abstract amodal rule. We address these possibilities in Experiment 2.

Exp 2: Common labels and responses

The goal of the second experiment is to understand whether learners are able to exploit the similarities across modalities when using the *same* labels and responses for categories across modalities ("1", "2" vs. Exp 1: auditory: "1", "2"; visual: "3", "4"). With the same labels/responses, learners may be better able to exploit the high-level regularities (i.e., shared rule-based structure) across modalities. As a result, this small procedural change should give participants a better opportunity for facilitation using amodal category representations if they are able to form them.

Method

Participants Participants were 66 undergraduate students from the University of New Hampshire, ages 18-22 (51 F, 15 M), who participated for partial course credit. Participants completed all tasks online through the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). This differs somewhat from Experiment 1 which was conducted in person, but prior work shows that in-person auditory and visual category learning results are overwhelmingly similar (Roark et al., 2021, 2022). Participants completed headphone checks to ensure they were wearing headphones (Milne et al., 2020).

Stimuli The stimuli were identical to Experiment 1.

Procedure The procedure was very similar to Experiment 1. Rather than seeing unique category labels for auditory and visual categories as in Experiment 1 (auditory: "1", "2"; visual: "3", "4"), which could also be linked to distinct motor responses (i.e., different button presses), in Experiment 2, all participants saw identical category labels for auditory and visual categories ("1", "2") for which participants presumably used the same motor responses.

Results

As in Experiment 1, auditory and visual performance was not significantly different when learned in separate blocks. This was true for both training ($t(34) = 0.52$, $p = .61$, $d = 0.088$) and generalization test (which was performed in separate blocks for both modalities after both training tasks; $t(34) = -0.24$, $p = .81$, $d = -0.040$).

Unlike Experiment 1, when participants used the *same labels/responses* for auditory and visual categories, performance was not significantly different across modalities (Figure 4; training: $F(1, 64) = 0.00094$, $p = .98$, $\eta_G^2 = 0.0000032$; test: $F(1, 64) = 1.16$, $p = .29$, $\eta_G^2 = 0.0030$) and did not differ across blocked or interleaved training (training: $F(1, 64) = 1.54$, $p = .22$, $\eta_G^2 = 0.018$; test: $F(1, 64) = 0.50$, $p = .48$, $\eta_G^2 = 0.0070$), demonstrating independence across modalities. There were no interactions between modality and training type (training: $F(1, 64) = 0.76$, $p = .39$, $\eta_G^2 = 0.0030$; test: $F(1, 64) = 0.50$, $p = .48$, $\eta_G^2 = 0.001$).

We next examined whether learners in the blocked training condition carried over any knowledge from one modality to the other modality – either facilitating or competing representations across modalities in the generalization test. Just as in Experiment 1, there were no significant differences in test performance when tasks were completed first and second (auditory: $t(33) = -0.22$, $p = .83$, $d = -0.074$; visual: $t(33) = 1.08$, $p = .29$, $d = 0.37$), demonstrating independence across modalities.

Overall, just as in Experiment 1, these results support a modality-specific framework of category learning. Further, when controlling for differences in cognitive demand induced by using different category labels/responses across modalities, the results from Experiment 2 suggest that rather than competing for resources or attention during learning, learners independently learned auditory and visual information. Further, even when supported by identical category labels/responses, learners were unable to exploit the shared domain-general rule to learn amodal representations shared across modalities. As such, there is neither a competition cost, nor a facilitation benefit in learning auditory and visual categories simultaneously. Instead, learners build independent, modality-specific category representations.

Discussion

Together, these results support a modality-specific framework of categorization and category learning. More specifically, these results primarily support independence across modalities such that both auditory and visual categories were not learned differently under separate, blocked conditions compared to interleaved learning.

Critically, even though the auditory and visual categories shared the same high-level regularities and could theoretically be learned using the same domain-general rule, learners did not exploit these regularities to facilitate learning across sensory modalities. As such, these results do not support the creation of shared amodal representations across auditory and visual modalities. Shared amodal representations were not formed even when participants used the same category label and response for both modalities. Future work is needed to assess whether learners can exploit shared structures within modalities or if category learning may be both stimulus- and modality-specific.

In Experiment 1, when auditory and visual categories required unique labels (auditory: “1”, “2”; visual: “3”, “4”), we observed competition effects across modalities during training. In the generalization test, when modalities were presented in different blocks for both conditions, there were no competition effects. These effects appear to be due to the additional cognitive challenge of switching back and forth between labels/response keys on different trials or generally switching between auditory and visual tasks as no competition was observed when categories required the same labels (auditory and visual: “1”, “2”) in Experiment 2. As a result, we interpret these competition effects as a byproduct of the task switching element of the design of the interleaved condition of Experiment 1, rather than competition between the modalities. That is not to say that the interleaved condition of Experiment 2 cannot still be conceptualized as task switching (i.e., between auditory and visual tasks), but switching may be less intrusive when one does not need to frequently switch between different response buttons. The remaining results support the conclusion of modality-specific representations that are independent across modalities.

Experiment 1 and 2 also differed in their medium – Experiment 1 was completed in person and Experiment 2 was completed online. While previous work has found many similarities between online and in-person auditory and visual category learning (Roark et al., 2021), it is also important to note that accuracy seems somewhat lower in our online-based experiment ($M = 64\%$) compared to the in-person experiment ($M = 71\%$). As there were other differences between the experiment that were theoretically driven (i.e., different vs. same buttons), we have primarily interpreted differences in patterns to be due to these manipulations. However, in future studies, it will be necessary to rule out potential alternative explanations by controlling for experimental medium.

Overall, these results add to the growing body of evidence that call for more direct consideration of the role of modality in models of categorization and category learning (Newell et al., 2023; Roark et al., 2021; Roark, Plaut, et al., 2022; Roark & Holt, 2019). We argue that it is necessary to look beyond vision to fully understand the mechanisms and nature of representations formed during category learning. More direct comparisons are needed to identify the specific ways in which auditory and visual learning rely on shared or specific cognitive processes. The current results emphasize that the common implicit assumption that models of categorization and category learning that primarily rely on experimental evidence from the visual modality are models of *domain-general* perceptual categorization and category learning should be reassessed and tested directly.

Further, it is also important to acknowledge that while we often experimentally isolate perception and learning within a single modality, natural real-world perception and learning are most often multimodal. Indeed, simply presenting multisensory information can alter perception and improve learning in important ways (Mathias & von Kriegstein, 2023; Murray & Shams, 2023; Shams & Seitz, 2008). While the current results focus on auditory and visual learning, it will be important for future work to continue moving towards more naturalistic, multimodal perception and learning environments.

These results also contribute to the longstanding debate in cognitive science about whether concepts and categories are generally represented via amodal or modality-specific representations (Kaup et al., 2023; Machery, 2016; Reed, 2016). Regarding learning, previous research demonstrated important modality-specific learning mechanisms involved in implicit statistical learning of auditory and visual sequence regularities (Conway & Christiansen, 2005, 2006; Frost et al., 2015; Mitchel & Weiss, 2011; Seitz et al., 2007). Here, we found similar modality-specific mechanisms in explicit supervised category learning. There is continued need to resolve this philosophical and theoretical debate about the nature of representations of human knowledge, such as those that are constructed during category learning.

In summary, our results support a modality-specific perspective on category learning. These results have important implications for understanding learning of real-world categories, which are often multimodal, and highlights the importance of considering the role of modality in models of categorization and category learning.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.

Bimbard, C., Sit, T. P. H., Lebedeva, A., Reddy, C. B., Harris, K. D., & Carandini, M. (2023). Behavioral origin of sound-evoked activity in mouse visual cortex. *Nature Neuroscience*, *26*(2), 251–258.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*(4), 1096–1108.

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

Conway, M. C., & Christiansen, M. H. (2006). Statistical learning within and between modalities. *Psychological Science*, *17*(10), 905–912.

Crossley, M. J., Filoteo, V., Maddox, T., & Ashby, F. G. (2023). Unique motor plans facilitate learning during task switching, but at the expense of greater switch costs. *PsyArXiv.*

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity : the paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125.

Garner, W. R. (1974). *The Processing of Information and Structure*. Erlbaum.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285.

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.

Kaup, B., Ulrich, R., Bausenhart, K. M., Bryce, D., Butz, M. V., Dignath, D., Dudschig, C., Franz, V. H., Friedrich, C., Gawrilow, C., Heller, J., Huff, M., Hütter, M., Janczyk, M., Leuthold, H., Mallot, H., Nürk, H.-C., Ramscar, M., Said, N., … Wong, H. Y. (2023). Modal and amodal cognition: an overarching principle in various domains of psychology. *Psychological Research*, 1–31.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—a review. *Psychological Bulletin*, *136*(5), 849–874.

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*, *23*(4), 1090–1095.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1–3), 59–70.

Mathias, B., & von Kriegstein, K. (2023). Enriched learning: behavior, brain, and computation. *Trends in Cognitive Sciences, 27*(1), 81–97.

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*, 12-Jan.

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.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Murray, C. A., & Shams, L. (2023). Crossmodal interactions in human learning and memory. *Frontiers in Human Neuroscience, 17*, 1181760.

Newell, F. N., McKenna, E., Seveso, M. A., Devine, I., Alahmad, F., Hirst, R. J., & O'Dowd, A. (2023). Multisensory perception constrains the formation of object categories: a review of evidence from sensory-driven and predictive processes on categorical decisions. *Philosophical Transactions of the Royal Society B*, *378*(1886), 20220342.

Reed, S. K. (2016). A taxonomic analysis of abstraction. *Perspectives on Psychological Science*, *11*(6), 817–837.

Roark, C. L., & Holt, L. L. (2019). Perceptual dimensions influence auditory category learning. *Attention, Perception, and Psychophysics*, *81*(4), 912–926.

Roark, C. L., Lescht, E., Wray, A. H., & Chandrasekaran, B. (2023). Auditory and visual category learning in children and adults. *Developmental Psychology*, *59*(5), 963–975.

Roark, C. L., Plaut, D. C., & Holt, L. L. (2022). A neural network model of the effect of prior experience with regularities on subsequent category learning. *Cognition*, *222*(104997), 104997.

Roark, C. L., Smayda, K. E., & Chandrasekaran, B. (2022). Auditory and visual category learning in musicians and nonmusicians. *Journal of Experimental Psychology: General*, *151*(3), 739–748.

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.

Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, *106*(34), 14611–14616.

Seitz, A. R., Kim, R., Wassenhove, V. van, & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36*(10), 1445–1453.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417.

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

Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *PLoS Biology*, *5*(3), e56.

Woolley, S. M. N., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, *8*(10), 1371–1379.