Title

Simultaneous Encoding Facilitates Multisensory Category Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Amiri, Hooman
Roark, Casey L

Publication Date
2026

Simultaneous Encoding Facilitates Multisensory Category Learning

Hooman Amiri (Hooman.Amiri@unh.edu) & Casey. L. Roark (Casey.Roark@unh.edu)
Department of Psychology, University of New Hampshire

Abstract

Category learning in the real world often relies on information from multiples senses, yet most studies of learning focus on a single modality. It is imperative to study multisensory learning not only to capture processes involved in naturalistic learning, but also because multisensory information may facilitate learning. In the current study, we examined how learning of audiovisual categories is impacted by presenting information simultaneously or separately across modalities. Overall, simultaneous presentation led to more successful performance, even when information was only presented in one modality. Simultaneous presentation also encouraged learners to attend more to the auditory modality. These findings suggest that simultaneous presentation of category information across different modalities can facilitate learning and that this may be related to how learners attend to information during learning. More broadly, these results demonstrate the necessity of investigating multisensory learning to reveal mechanisms supporting learning in more naturalistic multisensory contexts.

Keywords: multisensory; category learning; categorization; generalization; audiovisual

Introduction

Category learning is a core cognitive process by which individuals learn to organize stimuli into meaningful groups, allowing for generalization from known categories to new instances (Ashby & Maddox, 2005; Murphy 2002; Smith & Minda, 2000). When we learn to categorize objects in the environment, we usually do not rely on just one sensory modality. Our knowledge about cats, for example, involves visual appearance, auditory vocalizations, as well as tactile and olfactory experiences. People rely on redundant multisensory cues (e.g., acoustic and visual information cuing a /b/ phoneme) and complementary multisensory cues (e.g., visual appearance of a cat and "meow" sound) to learn categories across the lifespan (Newell et al., 2023). Despite the importance of multisensory information in everyday categorization, relatively little is known about how people *learn* multisensory categories as most studies of category learning have focused on learning from a single sense, typically vision (Ashby & Maddox, 2011; Minda et al., 2024; Richler & Palmeri, 2014).

Beyond relevance to real-world behavior, understanding multisensory category learning is also critical because there are reasons to expect that presenting information in multiple modalities may *facilitate* learning (Mathias & von Kriegstein, 2023; Murray & Shams, 2023; Shams & Seitz, 2008). Research in recognition memory demonstrates that merely presenting a congruent, redundant sound (e.g., "meow") alongside an image (e.g., cat) facilitates later recognition memory of that image (Heikkilä et al., 2015; Soto-Faraco & Spence, 2025). Even pairing arbitrary shapes with redundant sounds can enhance memory for those shapes (Brunel et al., 2013; Lauzon et al., 2022; Thelen et al., 2015). Relatively low-level perceptual detection and discrimination judgments are also facilitated by the simultaneous presentation of information in multiple modalities (Lippert et al., 2007; McDonald et al., 2000; Noesselt et al., 2008; Seitz et al., 2006). This research suggests that presenting information in multiple modalities may facilitate category learning.

Several studies have found that presenting redundant multisensory information benefits category learning compared to unisensory information alone. For example, incidental learning of visual categories (e.g., cartoon frogs) is enhanced when a category-specific auditory cue (e.g., unique 'croak' sound) is presented during training (Broadbent et al., 2020, 2019). Similarly, learning of novel auditory speech categories (e.g., Mandarin tones) is enhanced when a category-specific visual cue (e.g., line or hand gesture matching the pitch direction) is presented during training (Hannah et al., 2017; Zhen et al., 2019). Multisensory presentation also improves generalization to novel *unisensory* category exemplars (Broadbent et al., 2020), suggesting that this benefit is not simply due to the specific information encountered during learning.

Critically, in these studies, learning performance was enhanced when information in a second modality was present compared to when it was *absent* and only unisensory information was presented. As a result, improved performance may stem from either information being *multisensory* or from there being *more redundant information* available during encoding (Broadbent et al., 2020; Hendriks et al., 2023). Thus, there is a need to disentangle multisensory information from the amount of information available during encoding. In the current study, we disentangle these explanations by examining audiovisual category learning with the same multisensory information presented *simultaneously* or *separately* within the same trial. Conditions only differ in whether auditory and visual information is encoded at the same time or at separate times. If *multisensory* information facilitates learning (regardless of when it is presented) by learners having access to redundant cues in different modalities to support categorization decisions, learning should not be different across presentation types. However, if having access to redundant cues *at the*

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

moment of encoding facilitates learning, learning should be superior when auditory and visual information are encoded simultaneously rather than separately.

Importantly, how learners attend to stimulus information to make categorization decisions may differ when stimuli are presented simultaneously or separately. When information is redundant across modalities and presented simultaneously, people often rely more on the *visual* modality (Colavita, 1974; Colavita & Weisberg, 1979). For example, in speeded detection tasks where participants are asked to respond to the appearance of a light or a tone, they often fail to respond to the tone when both are presented simultaneously (Colavita, 1974; Colavita & Weisberg, 1979). This suggests that simultaneously presented auditory and visual information may compete for attention, which may result in learners shifting attention toward one modality (Lavie et al., 2014; Raveh & Lavie, 2015). In the current study, we will examine how learners attend to auditory and visual information to make categorization decisions in a test phase. If learners primarily rely on one modality to learn categories with redundant auditory and visual cues, when given only unisensory information, their performance should be better in one modality than the other. Similarly, if learners primarily rely on one modality, when they counter incongruent auditory and visual information, their responses should more closely align with one modality.

Finally, based on prior work, attention may be influenced by the time available to process the stimuli, with shorter presentation times encouraging attention to the auditory modality (Robinson & Sloutsky, 2019) as auditory processing is more temporally precise than visual processing (Penney et al., 2000; Shamma, 2001). To account for these possible differences in attention based on processing time, we compare learning of both simultaneous and separate presentation with relatively short or long total stimulus durations. We predict that shorter processing times may encourage learners to attend more to auditory than visual information.

To summarize, in the present study we examine how learning of redundant audiovisual categories is impacted by the simultaneous (vs. separate) presentation of information across modalities. This study will disentangle two potential sources of facilitation observed in prior studies – multisensory information and the amount of information available during encoding. We also examine how presentation type impacts how learners attend to information across modalities and whether this depends on the amount of time learners have to process information.

Method

We pre-registered the experiment design and report all planned confirmatory analyses (https://osf.io/6vsc9/overview?view_only=c66491d308f54d079e84ecabef96b1fb).

Participants

Participants were 132 undergraduate students from the University of New Hampshire ages 18-26 (20 Male, 112 Female) who completed the study in person for partial course credit. Participants all had normal or corrected-to-normal vision and hearing. All participants provided informed consent prior to participation. Participants were randomly assigned to one of four conditions varying in presentation type (Simultaneous, Separate) and total stimulus duration (Short [1 sec], Long [2 sec]): Simultaneous-Long (N = 34), Simultaneous-Short (N =31), Separate-Long (N = 33), Separate-Short (N =34). According to the power analysis outlined in the pre-registration, we recruited a total of 132 participants after exclusion of 56 participants due to failure of auditory and visual attention checks (< 85% accuracy). Although this exclusion rate was high, the criterion was necessary to ensure that participants were attending to both modalities and were meaningfully engaged with the task.

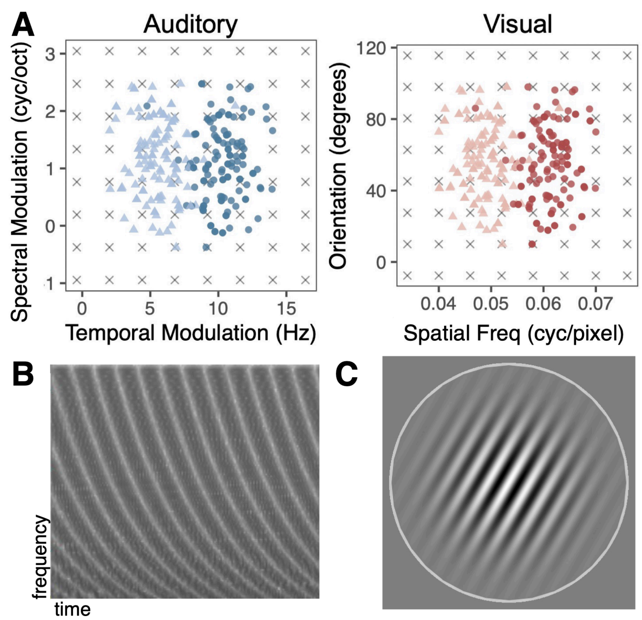

Figure 1. Stimulus distributions and example stimuli

Stimuli

All participants learned two multisensory categories comprised of redundant visual and auditory stimuli. The categories and stimuli for training were identical to previous unisensory learning studies (Roark, 2024; Roark et al., 2021). Category distributions were created sampling from a bivariate Gaussian distribution in normalized space and were then transformed into auditory (Figure 1A) and visual (Figure 1B) stimulus spaces. Categories could be distinguished with a unidimensional rule using information in either modality.

Auditory stimuli were complex nonspeech ripple sounds varying in temporal (Hz) and spectral modulation (cycles per octave); Figure 1C). Visual stimuli were 8cm x 8cm Gabor patches (Figure 1D), presented on a grey background and varying in spatial frequency (cycles per pixel) and orientation (degrees). There were 100 visual and 100 auditory stimuli (50 per category) presented in the training phase. A separate set of auditory and visual stimuli were created for the generalization test, sampling an 8x8 grid in the same spaces as the training stimuli (Xs in Figure 1A/B; 64 stimuli).

The stimuli could be optimally sorted into two categories based either on the spatial frequency dimension of the Gabor patches *or* the temporal modulation rate of the ripple sounds. That is, auditory and visual information are redundant. Specifically, optimal performance could be obtained by categorizing Gabor patches with frequencies greater than 0.05 cycles per pixel as Category A and those less than 0.05 cycles per pixel as Category B *or* categorizing ripple sounds with temporal modulation rates faster than 7.5 Hz as Category A and those slower than 7.5 Hz as Category B.

Procedure

The experiment was run using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). The procedure included both a training (200 trials across 8 blocks of 25 trials) and test phase (128 trials split across 4 blocks). Auditory stimuli were presented through Sennheiser over-ear headphones. Auditory and visual stimuli were presented at the same time in the Simultaneous condition (Figure 2A) and separately in the Separate condition (with modality order counterbalanced across trials; Figure 2B). The total duration of stimulus presentation varied across conditions being relatively Short (Separate: 0.5 sec + 0.5 sec; Simultaneous: 1 sec) or Long (Separate: 1 sec + 1 sec; Simultaneous: 2 sec). To avoid the possible perception of the separately presented auditory and visual stimuli as overlapping due to persistence of stimulus information during the presentation of the second stimulus, Separate trials always include a 150 ms interstimulus interval (Bowling & Lovegrove, 1980; Efron, 1970a, 1970b).

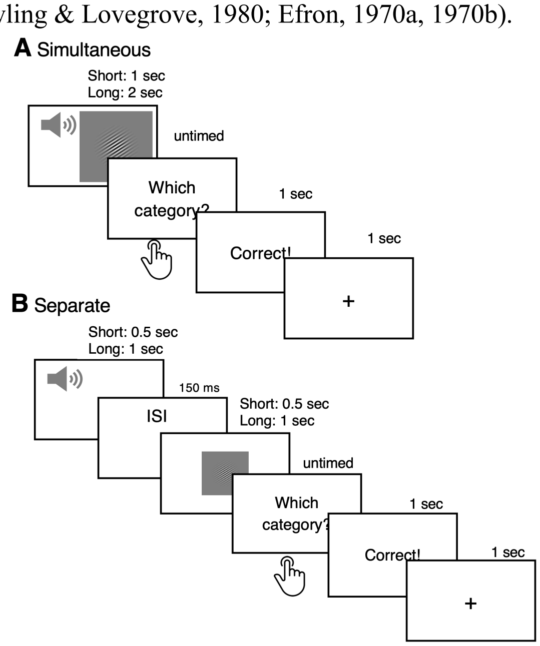

Figure 2. Trial Procedures across Conditions. The order of auditory and visual stimuli in the Separate condition was counterbalanced.

After stimulus presentation, participants were prompted to respond which category the audiovisual stimuli belonged to by pressing 1 or 2 keys on the keyboard. The assignment of keys to different categories was counterbalanced across participants. Participants had an unlimited amount of time to respond. Immediately after their response, participants received feedback (Correct or Incorrect) for 1 sec, followed by a 1 sec inter-trial interval. Attention check trials occurred five times each in every training block and required pressing the space bar within 2 sec after hearing a 1 sec 1000 Hz pure tone or seeing a visual cue to press the spacebar.

In the test phase, participants categorized auditory-only (32 trials), visual-only (32 trials), and audiovisual stimuli (64 trials) *without feedback*. Unisensory test stimuli demonstrate how learners cope with having only partial information relative to information received during training and can reveal how learners attend to modality information during learning. Half of the audiovisual trials were *congruent* (auditory and visual signal same category identity) and the other half were *incongruent* (auditory and visual signal different category identities). Incongruent stimuli can reveal how learners attend to modalities to resolve incongruent information. Further, we counterbalanced all test trial types (separate/simultaneous and short/long presentation) to not unfairly advantage any condition. Per the pre-registration, we analyzed data collapsed across trial types and present significant differences across audiovisual-separate and simultaneous test trial types as an exploratory analysis.

Results

Training phase

Mixed-model ANOVAs were conducted to examine the effects of presentation type (Separate vs. Simultaneous; between-subjects), total duration of stimulus presentation (1 sec vs. 2 sec; between-subjects), and block (1-8; within-subjects) on accuracy during training (Figure 3).

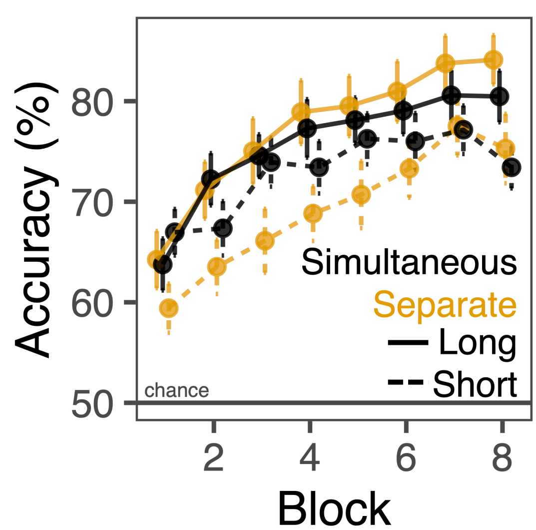

Figure 3. Performance in Training Phase
Note. Error bars represent *SEM*.

Across conditions, participants showed evidence of learning as performance increased across blocks ($F(7, 896) = 35.3$, $p < .0001$, $\eta_g^2 = 0.098$). Training performance did not differ when auditory and visual information was presented Simultaneously (M = 74%) or Separately (M = 73%; $F(1, 128) = 0.29$, $p = .59$, $\eta_g^2 = 0.001$). Training performance was significantly better when stimuli were presented for relatively long (M = 76%) compared to short durations (M = 71%; $F(1, 128) = 6.14$, $p = .015$, $\eta_g^2 = 0.098$). There were no significant interactions ($ps > .11$).

Test phase

Critically, to understand whether simultaneously presenting information from different sensory modalities benefits learning and impacts how people attend to modality information, it is necessary to directly compare performance on *identical stimuli* in the test phase. We first examined performance on congruent audiovisual trials and when information was presented in only one modality (Figure 4).

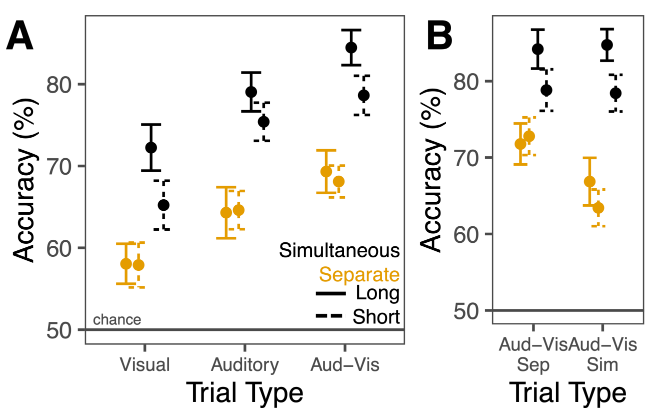

Figure 4. Performance in Test Phase
Note. Error bars represent *SEM*.

Congruent Audiovisual and Unisensory Stimuli. A mixed-model ANOVA was conducted to examine the effects of presentation type during training (Separate vs. Simultaneous), total duration of stimulus presentation (1 sec vs. 2 sec), and trial type (auditory-only, visual-only, audiovisual) on test accuracy (Figure 4A). Participants who experienced auditory and visual stimuli Simultaneously during training (M = 76%) performed better on the test overall compared to those who were trained with Separate presentation (M = 64%; $F(1,128) = 33.9$, $p < .0001$, $\eta_g^2 = 0.15$).

Performance differed across trial types ($F(2, 256) = 45.4$, $p < .0001$, $\eta_g^2 = 0.10$) such that accuracy was highest when both auditory and visual information was available (M = 75%), followed by auditory-only (M = 71%), and lowest when only visual information was available (M = 63%; all pairwise comparisons were significant, $ps < .028$). There was no significant interaction between presentation and trial type ($F(2, 256) = 0.45$, $p = .64$, $\eta_g^2 = 0.001$). Test accuracy did not differ across training duration ($F(1, 128) = 1.97$, $p = .16$, $\eta_g^2 = 0.010$) and there were no other interactions ($ps > .21$).

Exploratory Analysis: Generalization to different audiovisual presentation types. Next, as a planned exploratory analysis, we examined whether Separate and Simultaneous performance on *audiovisual* test stimuli differed based on whether the test stimuli were presented separately or simultaneously (Figure 4B). This comparison assesses whether learning generalizes to a different presentation type than the one experienced during training.

A mixed-model ANOVA was conducted to examine the effects of presentation type, duration, and trial type with two types of audiovisual stimuli (auditory-only, visual-only, audiovisual-separate, audiovisual-simultaneous) on test accuracy. Importantly, when considering audiovisual trial types separately, there was a significant interaction between presentation and trial type ($F(3, 384) = 2.99$, $p = .031$, $\eta_g^2 = 0.0080$).

For participants who experienced auditory and visual stimuli Simultaneously during training, accuracy was not significantly different for audiovisual test trials that matched their training type (simultaneous: M = 82%) or were different (separate: M = 82%; $p = .97$). Performance on auditory-only trials (M = 77%) was not significantly different from either type of audiovisual trials ($ps > .25$). However, performance on visual-only trials (M = 69%) was worse than both types of audiovisual trials ($ps < .0001$) and auditory-only trials ($p = .0048$). When training involved simultaneous presentation of auditory and visual information, participants coped well with missing visual information, but performance suffered when auditory information was missing. These results suggest that Simultaneous encoding encourages attention to the auditory modality and successfully generalizes to separate presentation of audiovisual information.

In contrast, for participants who experienced information Separately during training, performance on auditory-only (M = 64%; $p = .017$) and visual-only trials (M = 58%; $p < .0001$) was worse than trials where audiovisual information was presented separately (M = 72%). Performance was worse on visual-only ($p = .029$), but not auditory-only ($p = .86$) trials than trials where audiovisual information was presented simultaneously (M = 65%). This suggests that Separate presentation encourages attention to information in both modalities as performance suffered when either modality was missing. Further, accuracy was significantly better for audiovisual test trials that matched their training type (separate: M = 72%) than those that were different (simultaneous: M = 65%; $p = .047$, $d = 0.45$), suggesting that Separate encoding does not completely generalize to simultaneous presentation.

Incongruent Audiovisual Trials. Finally, we examined how learners attended to modality information when stimuli were incongruent across modalities (Figure 5). For these stimuli there is no 'correct' answer, but the proportion of trials where a participants' response aligns with a specific modality can indicate how they attended to information on these trials. We examined the proportion of responses aligned with the auditory modality such that values > 0.5 indicate more attention to auditory information, values < 0.5 indicate more attention to visual information, and values around 0.5 indicate similar attention to both modalities. We first examined responses in the four conditions using separate one-sample t-tests that compare the proportion of auditory-aligned responses to 0.5 and then compared responses across conditions using a mixed-model ANOVA with presentation type during training (Separate vs. Simultaneous) and total duration of stimulus presentation (1 sec vs. 2 sec).

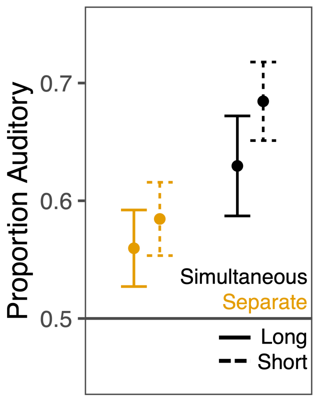

Figure 5. Performance on Incongruent Audiovisual Trials
Note. Error bars represent *SEM*.

Participants who experienced auditory and visual stimuli Simultaneously during training attended more to the auditory modality (Simultaneous-Short: $M = 0.68$, $SD = 0.19$; $t(30) = 5.53$, $p < .0001$, $d = 0.99$; Simultaneous-Long: $M = 0.63$, $SD = 0.25$; $t(33) = 3.05$, $p = .0045$, $d = 0.52$). Participants who experienced auditory and visual stimuli Separately and for a *short duration* during training attended slightly more to the auditory modality ($M = 0.58$, $SD = 0.18$; $t(33) = 2.72$, $p = .010$, $d = .047$) and those who experienced Separate stimuli for *long durations* did not specifically attend to either modality ($M = 0.56$, $SD = 0.19$; $t(32) = 1.84$, $p = .076$, $d = 0.32$). Importantly, attention to the auditory modality was stronger after Simultaneous ($M = 0.66$) than Separate training ($M = 0.57$; $F(1, 128) = 5.783$, $p = .018$, $\eta_g^2 = 0.043$). There were no differences responses to incongruent stimuli based on duration ($F(1,128) = 1.28$, $p = .26$, $\eta_g^2 = 0.148$, $p = .26$, $\eta_g^2 = 0.01$) and no interaction between presentation type and duration ($F(1,128) = 0.18$, $p = .67$, $\eta_g^2 = 0.001$). These results suggest that simultaneously encountering auditory and visual information during learning may more strongly encourage attention to the auditory modality than when information is encountered separately.

Discussion

We found that Simultaneous presentation of redundant auditory and visual category information *facilitated learning* compared to Separate presentation of the same information. Simultaneous learning generalized more successfully to encountering audiovisual information separately and was more successful when only one modality was available (especially the auditory modality). These results suggest that multisensory learning may not be enhanced by *multisensory* information or redundant information per se, but rather from there being redundant information available *during encoding*.

Experiencing co-occurring auditory and visual cues may better resemble real-world learning conditions, where people rarely rely on a single sense (Shams & Seitz, 2008). The brain is highly sensitive to temporal synchrony across modalities, with enhanced neural responses when information is presented simultaneously in multiple modalities (Stein & Stanford, 2008). Multisensory integration mechanisms support low-level perceptual performance (i.e., detection and discrimination) by producing stronger single-neuron responses to multisensory (i.e., AV) information than the linear combination of both modalities (i.e., A + V; Meredith & Stein, 1983; Stanford et al., 2005; Rowland & Stein, 2007). These 'superadditive' neural responses are highly sensitive to temporal synchrony of the signals, becoming weaker as modalities become more offset from one another (Marchant et al., 2012; Meredith et al., 1987; Noesselt et al., 2008). This suggests that the benefits observed in the current study may stem from enhancement of perception of multisensory information due to the precise temporal alignment of learning in the Simultaneous condition.

Further, our results align with and extend prior work in learning and memory demonstrating benefits to multisensory learning even after controlling for the overall information present. While prior studies have found benefits of multisensory learning compared to *unisensory* learning, it is unclear whether benefits were due to multisensory information per se or the overall amount of information present. For example, several studies have found that more redundant information available during encoding (vs. less) improves learning and memory regardless of whether that information comes from one modality or multiple modalities (Broadbent et al., 2020; Hendriks et al., 2023). Thus, these studies propose an alternative mechanism of potential benefits of multisensory learning – more redundant information present during encoding results in more varied representations that can be activated more successfully later even when some information is missing (i.e., unisensory test stimuli; Hendriks et al., 2023; Lauzon et al., 2022; Thelen et al., 2015). Our results generally support this perspective and extend these findings to also control for the overall amount of information present *within a trial* further demonstrating that learning benefits are specific to the amount of information present *during encoding*.

Varying the amount of information present during encoding also impacted how learners attended to information in different modalities during learning. Our results suggest that in the case of category learning, Simultaneous presentation may encourage stronger attention to the auditory modality, preserving performance when only auditory information was available and using auditory information to resolve incongruencies between modalities. One possible explanation for these differences in attention is that stimuli were presented for a relatively short period of time (1-2 seconds). The auditory modality is highly sensitive to temporal information (Penney et al., 2000; Shamma, 2001) and people tend to use auditory information more when information is presented for relative short durations (Robinson & Sloutsky, 2010). In the current study, presentation was quite short even in the long duration conditions (2 seconds). Most studies of visual category learning present stimuli until a response is made (Ashby & Maddox, 2011; Richler & Palmeri, 2014). While this provides ample time to process all parts of a visual stimulus, it is challenging to use a comparable technique for auditory information and may not capture how people experience information in many real-world situations. These relatively short presentation times may have encouraged learners to rely less on the visual information as they were able to retain processing of short auditory stimuli while visual processing was impaired (Emberson et al., 2011; Wearden & Jones, 2020). While our findings may seem incongruent with prior work demonstrating visual dominance with redundant auditory and visual information (Colavita, 1974; Colavita & Weisberg, 1979), they are consistent with studies demonstrating that visual dominance is reduced when information is presented at shorter durations (Robinson & Sloutsky, 2019). These results highlight the importance of understanding how attention functions in multisensory environments and how this subsequently impacts learning.

In the current study, we found that multisensory category learning is impacted by the way information is presented during learning, with simultaneous presentation of redundant auditory and visual information facilitating learning compared to separate presentation of the same information. However, it remains unclear whether these differences reflect a benefit of simultaneous presentation, a cost of separate presentation, or a combination of both. Our results extend to theoretical perspectives on the possible benefits of multisensory learning suggesting that simultaneous encoding of information across modalities facilitates learning above and beyond presentation of redundant multisensory information. Our results also demonstrate that modalities compete for attention, negatively impacting category learning when information is presented separately and encouraging learners to attend to auditory information more strongly when information is presented simultaneously. In general, these results suggest that to understand how we learn real-world multisensory categories, it is critical to examine attention and learning in multisensory contexts. While the present study used novel category stimuli that allowed precise control over the amount and timing of multisensory information, naturalistic categories may contain less clearly redundant cues across modalities. Future work should examine whether similar benefits of simultaneous encoding emerge when learning more naturalistic multisensory categories. Even with artificial highly controlled stimuli, the current results challenge current approaches and theories of category learning which do not explicitly account for learning within or across different sensory modalities.

References

Ashby, F. G. & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

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

Bowling, A. & Lovegrove, W. (1980). The effect of stimulus duration on the persistence of gratings. *Perception & Psychophysics*, *27*(6), 574–578. https://doi.org/10.3758/bf03198688

Broadbent, H., Osborne, T., Mareschal, D. & Kirkham, N. Z. (2019). Withstanding the test of time: Multisensory cues improve the delayed retention of incidental learning. *Developmental Science*, *22*(1), e12726. https://doi.org/10.1111/desc.12726

Broadbent, Osborne, T., Mareschal, D. & Kirkham, N. (2020). Are two cues always better than one? The role of multiple intra-sensory cues compared to multi-cross-sensory cues in children's incidental category learning. *Cognition*, *199*, 104202. https://doi.org/10.1016/j.cognition.2020.104202

Brunel, L., Goldstone, R. L., Vallet, G., Riou, B. & Versace, R. (2013). When seeing a dog activates the bark: Multisensory generalization and distinctiveness effects. *Experimental Psychology*, *60*(2), 100–112. https://doi.org/10.1027/1618-3169/a000176

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. https://doi.org/10.3758/bf03203962

Colavita, F. B. & Weisberg, D. (1979). A further investigation of visual dominance. *Perception & Psychophysics*, *25*(4), 345–347. https://doi.org/10.3758/bf03198814

Efron, R. (1970a). Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics*, *8*(4), 231–234. https://doi.org/10.3758/bf03210211

Efron, R. (1970b). The minimum duration of a perception. *Neuropsychologia*, *8*(1), 57–63. https://doi.org/10.1016/0028-3932(70)90025-4

Emberson, L. L., Conway, C. M. & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, *64*(5), 1021–1040. https://doi.org/10.1080/17470218.2010.538972

Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J. & Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in mandarin tone perception in noise by native and non-native perceivers. *Frontiers in Psychology*, *8*, 2051. https://doi.org/10.3389/fpsyg.2017.02051

Heikkilä, J., Alho, K., Hyvönen, H. & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology*, *62*(2), 123–130. https://doi.org/10.1027/1618-3169/a000279

Hendriks, D., Verkoeijen, P. & Pecher, D. (2023). Encoding variability explains the multisensory benefit in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(6), 920–929. https://doi.org/10.1037/xlm0001305

Lauzon, S. Á., Abraham, A. E., Curcin, K., Butler, B. E. & Stevenson, R. A. (2022). The relationship between multisensory associative learning and multisensory integration. *Neuropsychologia*, *174*, 108336. https://doi.org/10.1016/j.neuropsychologia.2022.108336

Lavie, N., Beck, D. M. & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130205. https://doi.org/10.1098/rstb.2013.0205

Lippert, M., Logothetis, N. K. & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, *1173*, 102–109. https://doi.org/10.1016/j.brainres.2007.07.050

Mathias, B., & von Kriegstein, K. (2023). Enriched learning: Behavior, brain, and computation. *Trends in Cognitive Sciences*, *27*(1), 81–97. https://doi.org/10.1016/j.tics.2022.10.007

McDonald, J. J., Teder-Sälejärvi, W. A. & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908. https://doi.org/10.1038/35038085

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215–3229. https://doi.org/10.1523/jneurosci.07-10-03215.1987

Minda, J. P., Roark, C. L., Kalra, P. & Cruz, A. (2024). Single and multiple systems in categorization and category learning. *Nature Reviews Psychology*, *3*(8), 536–551. https://doi.org/10.1038/s44159-024-00336-7

Murphy, G. (2002). *The Big Book of Concepts*. https://doi.org/10.7551/mitpress/1602.001.0001

Murray, C. A. & Shams, L. (2023). Crossmodal interactions in human learning and memory. *Frontiers in Human Neuroscience*, *17*, 1181760. https://doi.org/10.3389/fnhum.2023.1181760

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: Biological Sciences*, *378*(1886), 20220342. https://doi.org/10.1098/rstb.2022.0342

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J. & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, *1220*, 157–163. https://doi.org/10.1016/j.brainres.2007.12.060

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1770–1787. https://doi.org/10.1037/0096-1523.26.6.1770

Raveh, D. & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*(2), 483–492. https://doi.org/10.3758/s13414-014-0776-2

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

Robinson, C. W. & Sloutsky, V. M. (2010). Effects of multimodal presentation and stimulus familiarity on auditory and visual processing. *Journal of Experimental Child Psychology*, *107*(3), 351–358. https://doi.org/10.1016/j.jecp.2010.04.006

Robinson, C. W. & Sloutsky, V. M. (2019). Two mechanisms underlying auditory dominance: Overshadowing and response competition. *Journal of Experimental Child Psychology*, *178*, 317–340. https://doi.org/10.1016/j.jecp.2018.10.001

Rowland, B. A., & Stein, B. E. (2007). Multisensory integration produces an initial response enhancement. *Frontiers in Integrative Neuroscience*, *1*, 4. https://doi.org/10.3389/neuro.07.004.2007

Seitz, A. R., Kim, R. & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*(14), 1422–1427. https://doi.org/10.1016/j.cub.2006.05.048

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, *5*(8), 340–348. https://doi.org/10.1016/s1364-6613(00)01704-6

Shams, L. & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Smith, J. D. & Minda, J. P. (2000). Thirty Categorization Results in Search of a Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Soto-Faraco, S. & Spence, C. (2025). Crossmodal semantics in memory: scoping review and meta-analyses of multisensory effects in short-term and episodic memory systems. *Psychological Bulletin*. https://doi.org/10.1037/bul0000483

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *The Journal of Neuroscience*, *25*(28), 6499–6508. https://doi.org/10.1523/jneurosci.5095-04.2005

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. https://doi.org/10.1038/nrn2331

Thelen, A., Talsma, D. & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148–160. https://doi.org/10.1016/j.cognition.2015.02.003

Wearden, J. H. & Jones, L. A. (2020). Judgements of the duration of auditory and visual stimuli. *Timing & Time Perception*, *9*(2), 199–224. https://doi.org/10.1163/22134468-bja10008

Zhen, A., Hedger, S. V., Heald, S., Goldin-Meadow, S. & Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition*, *187*, 178–187. https://doi.org/10.1016/j.cognition.2019.03.004