Cognition 238 (2023) 105542

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Evidence for an amodal domain-general object recognition ability

Jason K. Chow [a,*], Thomas J. Palmeri [a], Graham Pluck [b], Isabel Gauthier [a]

[a] Department of Psychology, Vanderbilt University, USA
[b] Faculty of Psychology, Chulalongkorn University, Thailand

ARTICLE INFO

Keywords:
Object recognition
Audition
Vision
Individual differences
Amodal

ABSTRACT

A general object recognition ability predicts performance across a variety of high-level visual tests, categories, and performance in haptic recognition. Does this ability extend to auditory recognition? Vision and haptics tap into similar representations of shape and texture. In contrast, features of auditory perception like pitch, timbre, or loudness do not readily translate into shape percepts related to edges, surfaces, or spatial arrangement of parts. We find that an auditory object recognition ability correlates highly with a visual object recognition ability after controlling for general intelligence, perceptual speed, low-level visual ability, and memory ability. Auditory object recognition was a stronger predictor of visual object recognition than all control measures across two experiments, even though those control variables were also tested visually. These results point towards a single high-level ability used in both vision and audition. Much work highlights how the integration of visual and auditory information is important in specific domains (e.g., speech, music), with evidence for some overlap of visual and auditory neural representations. Our results are the first to reveal a domain-general ability, *o*, that predicts object recognition performance in both visual and auditory tests. Because *o* is domain-general, it reveals mechanisms that apply across a wide range of situations, independent of experience and knowledge. As *o* is distinct from general intelligence, it is well positioned to potentially add predictive validity when explaining individual differences in a variety of tasks, above and beyond measures of common cognitive abilities like general intelligence and working memory.

There is increasing recognition that high-level perceptual abilities are distinct from general intelligence and are not well predicted by tests of cognitive skills (Richler et al., 2019; Richler, Wilmer, & Gauthier, 2017) that are the focus of many selection processes in society (Dunnette, 1966; Lubinski, 2000). High-level perception relies on the integration of numerous sensory dimensions into perceptual objects that are relatively invariant to presentation conditions, can be categorized and associated with non-perceptual information, and have behavioral significance (Palmeri & Gauthier, 2004). High-level perceptual skills are relevant to a variety of occupations and identifying individuals with high perceptual ability may improve prediction of performance and training (Sunday, Donnelly, & Gauthier, 2018). One challenge is to determine the structure of perceptual abilities, including whether there may be abilities common across modalities of perception. Fully characterizing the dimensions of individual differences in perception would require multivariate studies using a broad range of different tests. However, there is a paucity of psychometrically-sound measures for high-level perceptual abilities relative to measures for low-level perceptual abilities (e.g., contrast sensitivity, pitch discrimination; Pelli & Bex, 2013; Watson, Johnson, Lehman, Kelly, & Jensen, 1982). High-level perception is usually discussed at the level of an entire modality (e.g., vision; Albright, 2013) or even more specifically within a modality (e.g., scene perception, Henderson & Hollingworth, 1999). Here we seek evidence for the existence of a general high-level object recognition ability relevant across vision and audition.

Many individual differences studies in perception draw inferences based on correlations between pairs of similar tests. When performance is measured in tests with similar tasks (e.g., two matching tasks; Growns, Dunn, Mattijssen, Quigley-McBride, & Towler, 2022) or tests that tap similar domains (e.g., two speech perception tasks; Watson, Qiu, Chamberlain, & Li, 1996), any observed correlation is difficult to primarily ascribe to a domain-general ability, without the addition of more tests targeted at the domain-generality of abilities. We can understand more from proximal dissociations (lack of correlations between similar tests) and remote associations (correlations between dissimilar tests; Wilmer, 2008). We seek, for the first time, to measure the remote

* Corresponding author at: Department of Psychology, Vanderbilt University, 111 21st Avenue South, Nashville, TN 37240, USA.
E-mail address: jason.k.chow@vanderbilt.edu (J.K. Chow).

https://doi.org/10.1016/j.cognition.2023.105542
Received 16 November 2022; Received in revised form 26 June 2023; Accepted 27 June 2023

association between visual and auditory domain-general object recognition abilities.

Our approach builds on recent work on domain-general visual object recognition abilities. Performance on memory and matching tests with a range of novel objects was explained by a domain-general factor that was termed *o* (for object recognition; Richler et al., 2019; see also Ćepulić, Wilhelm, Sommer, & Hildebrandt, 2018). This study revealed that *o* was distinct from general intelligence and other cognitive and personality variables (see also Richler et al., 2017; Smithson, Chow, Chang, & Gauthier, 2023). *o* was essentially identical when measured using novel or familiar objects (Sunday, Tomarken, Cho, & Gauthier, 2022). As expected from a domain-general ability, it accounts for variability across a range of visual tests and domains (Chang & Gauthier, 2021; Gauthier & Fiestan, 2023; Sunday et al., 2018). While *o* was initially defined as a high-level visual ability, it also correlates with haptic object recognition (Chow, Palmeri, & Gauthier, 2022). As with the visual work, psychometrically-sound haptic tests varying in task requirements and using different categories of stimuli converged on a domain-general factor of haptic object recognition, termed o_h. In a study with 97 participants, haptic ability o_h shared about 25% of its variance with the corresponding visual ability o_v (Chow, Palmeri, & Gauthier, 2023).

Shape processing is important for both haptic and visual perception (e.g., Hummel, 2001; Klatzky, Lederman, & Metzger, 1985; Klatzky, Lederman, & Reed, 1987) and both behavioral and cognitive neuroscience studies suggest shared mechanisms underlying these two modalities (Amedi, 2002; Chow et al., 2023; Gaissert, Wallraven, Bülthoff, & Bulthoff, 2010; Lacey & Campbell, 2006; Lee Masson, Bulthé, op de Beeck, & Wallraven, 2016; Sathian & Lacey, 2022), which might, in part, explain the domain-general factor between haptics and vision. In contrast, the auditory modality is not generally used for shape processing for most people (see Thaler & Goodale, 2016, for a possible exception). Therefore, an important question for our understanding of the mechanisms that underlie *o* is whether they only support the recognition of objects through modalities that process shape and texture (like vision and haptics) or whether they can generalize to object recognition via audition. The novelty of our work lies in addressing the overlap of auditory and visual processing i) in complex object recognition, ii) using an individual differences approach, iii) by looking at domain-general abilities (rather than mere test performance), and iv) controlling for other sources of domain-general performance.

To assess the correlation between o_v and an analogous auditory ability (o_a), we first needed to create measures of high-level auditory abilities. Prior work estimating o_v and o_h focused on subordinate-level object perception, requiring discriminations between objects in the same basic-level category (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Conversely, tests measuring auditory individual differences have focused on low-level features, on the highly specialized domain of speech, or highly distinguishable familiar sounds (e.g., Kidd, Watson, & Gygi, 2007; Watson et al., 1982; Yang et al., 2020). In summary, asking whether a putative o_a relates to visual abilities first required the creation of new sensitive tests for subordinate-level auditory object perception.

In our first experiment, we introduce a new test of high-level auditory perception, targeting a specific category of auditory "objects" that are relatively confusable (bird songs). This initial attempt at measuring performance in a task tapping a high-level auditory ability demonstrated a robust correlation with visual object recognition. In our second experiment, we use three auditory tests to define a more general auditory ability. This allowed us to assess the correlation between higher order abilities o_a and o_v. Across both experiments, we controlled for general intelligence, perceptual speed (Ekstrom, French, Harman, & Derman, 1976), low-level visual perception (Kieseler, Dickstein, Krafian, Li, & Duchaine, 2022), and working memory (Wilhelm, Hildebrandt, & Oberauer, 2013) to assess evidence for a cross-modal visual/auditory object recognition ability that cannot be accounted for by domain-general abilities at a higher level (cognitive) or lower level (speed or discrimination of single visual dimensions).

1. Experiment 1

1.1. Methods

1.1.1. Participants

Participants were recruited online from Prolific. We used Bayesian optional stopping to determine our final sample size; we started with 30 participants and continued to collect data until the evidence for or against a correlation between every pair of object recognition tests within a modality exceeded the predetermined threshold of $BF_{+0} > 3$ or $BF_{+0} < 1/3$ (in line with the common interpretation for "substantial evidence" and using the same threshold as in our prior work with haptic abilities; Chow et al., 2023; Jeffreys, 1961). Initial data collection was split across two sessions to avoid fatigue and only participants who completed both sessions were used in the dataset. Additionally, we invited participants to a follow-up session with more tests for divergent validity purposes; any analyses that used these extra tests only used data from participants that completed all three sessions; reanalyzing the data with only participants that completed all three sessions does not drastically change results. Session 1 was approximately 20 min long, session 2 was approximately 25 min long, and session 3 was approximately 15 min long. Participants were paid equivalent to 8–9 dollars an hour over the three sessions. Details on these sessions and the tests are below. Ultimately, we recruited 90 participants in session 1, but the final sample size used for data analyses was 82 participants for session 1 and 2 (47 females, 35 males; mean age = 34.57 years, SD = 10.82 years; breakdown of exclusions in the Analysis section). Of those participants, 67 participants returned for the final third session (39 females, 28 males; mean age = 34.87 years, SD = 10.96 years). Data collection and procedures were approved by the Vanderbilt Institutional Review Board.

1.1.2. Test battery

We developed a new auditory object recognition test similar to our previous visual and haptic object recognition tests (Chow et al., 2022; Richler et al., 2019). For the new auditory test, we created test trials intended to capture a range of difficulty (the easiest trials would have most participants answer correctly and the hardest trials would have most participants at chance performance) and honed the test over multiple iterations in pilot testing. We created a set of trials with a range of difficulty initially based on experimenter judgment, with the intention that observed difficulty should range from nearly chance performance to nearly perfect performance across a sample. We collected data over three batches of 10, 15, and 20 participants to adjust the new auditory test. After each batch, we reordered trials based on the observed difficulty and replaced trials with negative item-total correlations or if all participants succeeded or failed on the trial. We achieved high reliability (>0.8) in the last batch. Any participants that participated in the creation of the test were excluded from participating in the main experiment.

The visual object recognition tests, used in prior work also began with relatively easy trials and progressively increased in difficulty to encourage participants. Every participant experienced the same trial order for each test to ensure that any order effects are shared across participants.

1.1.2.1. Auditory memory test (aMT-birds). In this new test, participants were tasked with remembering six target birdsongs. The birdsong stimuli were obtained from The Macaulay Library at the Cornell Lab of Ornithology. Six bird species were selected to be the targets. We chose species and birdsongs that were distinct to ensure the test was reasonable for novices. We also selected birdsongs from amongst 30 other species to act as distractors throughout the test. Each birdsong was

edited to be approximately 3–6 s, and no individual clip was ever repeated.

The test starts with a study phase: participants are introduced to each of the six birdsongs one at a time. For each birdsong, participants studied a 15-s audio clip of multiple examples of that birdsong concatenated together, these clips were not used as test clips. After participants had studied a target birdsong, they completed three easy test trials where the target was the birdsong recently studied. For each test trial, three audio clips (3–6 s each) were presented sequentially, and as each clip was presented, participants were prompted to respond with a key (*f*, *g*, and *h*, for first, second, or third, respectively) to indicate which clip was the target bird. The clip of the target bird was not the same exact clip as studied but instead a different clip of the same bird species, thus avoiding any recognition based on non-diagnostic features like audio quality. If the participant had not responded by the time the third clip finished, the audio clips repeated from the first. Participants could respond with any of the keys during any clip. No feedback was given. After the six target birdsongs were studied and tested, participants reviewed each of the original study clips sequentially to prepare for the test phase.

In the test phase, participants completed 30 test trials in which any of the six target birds could be the target in each trial. These test trials ranged in difficulty, initially being only slightly more difficult than the study phase test trials but increased in difficulty over trials; by the end, the distractor birdsongs were more similar to the targets than the earlier foils. This test was scored with percent correct, including the test trials during the study phase for a total of 48 trials, therefore, 33% correct was chance performance.

Participants also performed another auditory object recognition test as we intended to create an aggregate auditory ability akin to o_v. After data collection, we discovered a design flaw in that test that rendered it unusable. We omit results from this other auditory test in this experiment. More information on this test can be found in the supplementary materials. An improved version of that test, which involved the sounds of keyboard switches, will be used in Experiment 2.

1.1.2.2. Visual novel object memory test (vNOMT-Greebles). This visual task was first used in Richler et al. (2017). The Greebles (Gauthier & Tarr, 1997) are a set of 3D novel objects defined by their body shape and their protrusions (with parts unique for each object and in this set, positioned in an asymmetrical configuration). In the study phase, participants were introduced to each target object from three different views for 3000 ms each. After a target object was introduced, participants completed three easy test trials to select the recently studied target amongst distractors. This repeated until six target objects had been introduced. Afterwards, the participants reviewed all six targets at once for 20 s. After the study phase, participants completed the test phase where any target could appear on each trial, in which they needed to choose which of three objects was one of the six studied targets. The Greebles could be shown in different viewpoints and with visual noise overlaid. This test was scored with percent correct, including the test trials during the study phase for a total of 72 trials, therefore, 33% correct was chance performance.

1.1.2.3. Visual matching test (vMatch-Ziggerins). This task was first used in Richler et al. (2017). The Ziggerins (Wong, Palmeri, & Gauthier, 2009) are a set of 3D novel objects that are defined by the shape of a single vertical rod and two protrusions near the top and bottom. On each trial, the first Ziggerin was presented for 300 ms (150 ms in the second half of the test), followed by a visual mask of scrambled Ziggerins images for 500 ms, and finally by a second Ziggerin was presented. The trial ended when the participant responded either same or different with the *g* or *h* key, respectively. To further increase the difficulty of later trials, the objects were sometimes presented from different views, at different sizes, and in visual noise. Ziggerins were considered the same if they had the same configuration of identical parts, regardless of these irrelevant transformations. This test was scored with d'[1] over 124 trials, therefore, d' = 0 was chance performance.

1.1.2.4. Additional tests. In addition to the above object recognition tests, we collected data with online versions of the Matrix Matching test (Matrices; Pluck, 2019), two tests of Perceptual Speed (P-Speed): the Hidden Figures Test and the Identical Pictures Test (Ekstrom et al., 1976), and the Hanover Early Visual Assessment (HEVA; Kieseler et al., 2022). These tests were intended to provide divergent validity and provided possible alternative explanations for the relations between visual and auditory object recognition tests.

The Matrices test (Pluck, 2019; Fig. 1) is a short assessment of intelligence using visual matching with images to assess both fluid and crystallized intelligence. On each trial, an array of images with a missing piece was presented and participants were asked which of a set of options best fit the array. Half of the trials were visuospatial based, dependent on the colour and shape, and the other half were semantic based, dependent on the semantics of the array of images. If the correct option was selected, participants scored a point. The later semantic trials had two pieces missing and therefore required participants to respond with two different options, allowing participants to score up to two points.

The Perceptual Speed tests (Ekstrom et al., 1976; Fig. 1) required rapid matching of simple figures (Identical Pictures) and rapid identification of a geometric pattern hidden in more complex figures (Hidden Figures). For Identical Pictures, participants were tasked to select the exact same line drawing as the target drawing amongst a set of 5 similar options, with a new target presented every trial. For Hidden Figures, a specific target pattern was specified at the start of the test and participants had to decide whether or not it was present in a series of geometric

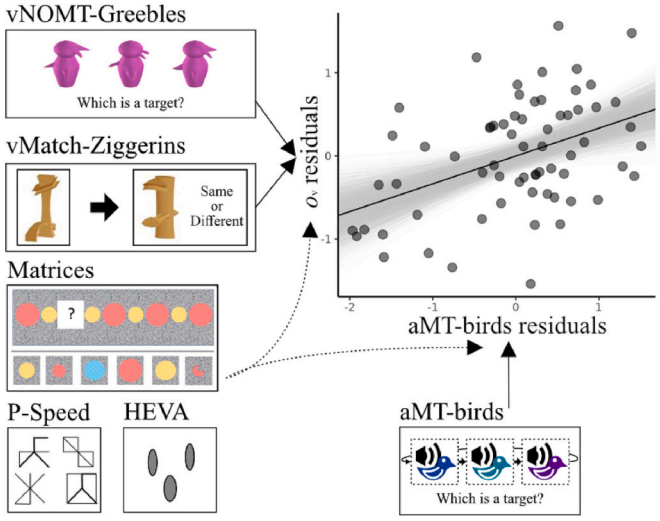

Fig. 1. Partial correlation between aMT-birds and o_v controlling for performance in Matrices, P-Speed, HEVA, and age. In the visual modality, we combined two tests with different task demands into a composite measure. To control for the effect general intelligence, perceptual speed, and low-level visual ability, we regress out scores on the Matrices, P-Speed, and HEVA. Each point in the scatterplot is a participant, the solid black line is a fit line with light gray lines representing different draws from the posterior.

[1] d' is calculated by the z-transform of the hit rate (proportion of trials answered "different" when the correct answer was "different") subtracted by the z-transform of the false alarm rate (proportion of trials answered "different" when the correct answer was "same").

patterns, presented one at a time. Both tests were timed tests and the score was the number correct minus the number incorrect within the time limit. The scores on each test were z-scored and averaged to form a P-Speed score.

The HEVA (Kieseler et al., 2022; Fig. 1) is a test of low-level visual processing. It required oddball judgments based on simple features such as line length or circle size. Each trial presented a set of three images placed on an imaginary circle to offset the placement of each image relative to another. The images only stayed on the screen for a brief period of time, but each trial would only end when a response was made. Participants were told what feature was relevant for the oddball judgment and the trials were blocked together based on what feature was being distinguished. This test was scored with percent correct, with chance level at 33% correct.

1.1.3. Procedure

Data collection was spread across three sessions online. In the first session, participants completed aMT-birds and vMatch-Ziggerins. After 24 h, in the second session, participants completed the other auditory test, vNOMT-Greebles, and Matrices. After two days, in the third session, participants completed the P-Speed tests and the HEVA. This split between the first two sessions was intended to reduce the effects of fatigue, but it also splits test format and modality across sessions to avoid correlations driven by within-session effects. Using a fixed test order across participants, common in individual differences research, avoids confounding order effects in individual differences but it does prevent us from evaluating any possible influence of order. Participants that completed the first session were invited to the second and third sessions, but not everyone returned to complete these later sessions.

1.1.4. Analyses

During tests without strict time-limits (all tests except the P-Speed tests), we added attention checks in the form of trials asking participants to click a specific option or to press a specific key (not one of the normal response keys). A single attention check trial was included in each test and would appear in the midst of all other test trials. They were targeted at excluding participants that were not paying attention to the task. We did not use data from any participant who failed two or more attention checks in a session, following the Prolific guidelines. Six participants were rejected in session one, one in session two, and none in session three. We also excluded one participant from the final dataset for scoring exceptionally low on the HEVA (38% accuracy, the next lowest participant being 58%). Repeating all analyses with this participant did not change the key results.

Bayesian analyses were performed using the BayesFactor package (Morey & Rouder, 2022) in R. For the correlational analyses, we expected small to moderate positive correlations, therefore we used directional hypotheses (H_+: positive correlation vs H_0: point null) using a default Jeffreys prior with a scaling parameter of $r = 1/3$. For the regression analysis, we used a default Jeffreys-Zellner-Siow prior (Wetzels & Wagenmakers, 2012) with scaling parameter of $r = 0.354$, again reflecting a weak prior for small to moderate effect sizes. To index uncertainty of our point estimates, we used the 95% highest posterior densities as credible intervals (95% CI).

For our object recognition tests, we do not assume that they uniquely tap into a single ability, therefore, to measure a general object recognition ability (in this case, o_v), we combine the scores from multiple tests (Smithson et al., 2023). For each participant, we calculated their o_v score by averaging the z-scored performance on each visual object recognition test. This aggregate indicator is less dependent on test-specific abilities and, instead, an indicator of domain-general ability.

1.2. Results

1.2.1. Correlations between object recognition measures

Descriptive statistics for each test are in Supplementary materials (Table S1). First, we assessed the bivariate correlations between our object recognition tests. Zero-order correlations are presented in Fig. 2. Estimates of reliability on our object recognition measures were acceptable (Fig. 1, diagonal). To estimate the true effect sizes of correlations, we report disattenuated correlations (r*; Table S1) to correct for measurement error, which is the observed correlations divided by the square root of the product of the reliability of each variable (Nunnally, 1994). This is useful to allow us to compare the effect sizes between correlations, accounting for attenuation from imperfect reliability. We found decisive evidence of a correlation between aMT and vNOMT-Greebles, $r = 0.41$, 95% CI [0.24, 0.58], $r^* = 0.49$, and between aMT-birds and vMatch-Ziggerins, $r = 0.40$, 95% CI [0.23, 0.57], $r^* = 0.48$ (all correlations and bayes factors are in Fig. 2). The similar magnitude of correlation between the aMT-birds and both visual tests suggests that the overlap between the aMT-birds and the visual tests is not driven primarily by shared task demands as aMT-birds shares more task demands with vNOMT-Greebles than vMatch-Ziggerins. We observed the expected correlation between our two visual object recognition tests. This leads us to further explore the relationship between aMT-birds and domain-general visual ability.

We tested whether performance on aMT-birds was related to general visual object recognition ability. As noted in the Method, we combined the two visual object recognition tests to estimate a higher-order domain-general latent variable (o_v).[2] This aggregate measure is more reliable than its components, o_v aggregate reliability[3] = 0.89. We found decisive evidence of a strong correlation between aMT-birds and o_v, $r = 0.51$, 95% CI [0.35, 0.66], $r^* = 0.56$, $BF_{+0} = 225{,}798.50$. This result suggests robust overlap in mechanisms used in the aMT-birds and visual object recognition ability.

1.2.2. Accounting for non-o variance

While the correlation between aMT-birds and visual ability was substantial, this correlation could be due to shared variance in domain-general mechanisms other than those specific to object recognition. While one can never measure all possible third variables, we chose to control for general intelligence, perceptual speed, and low-level perceptual abilities using Matrices, P-Speed, and HEVA. These measures were positively correlated amongst themselves and with the object recognition tests (Fig. 2). It is not surprising that any pair of these tasks should share some domain-general variance (Schneider & McGrew, 2012), be it cognitive or perceptual (which could be modal and/or amodal). Particularly, we found substantial evidence that aMT-birds correlated with Matrices and HEVA and we found inconclusive of a positive correlation with P-Speed. Further, o_v correlated with Matrices, $r = 0.43$, 95% CI [0.27, 0.61], $r^* = 0.59$, $BF_{+0} = 4399.04$; with P-Speed, $r = 0.24$, 95% CI [0.02, 0.42], $r^* = 0.26$, $BF_{+0} = 3.89$; and with HEVA, $r = 0.24$, 95% CI [0.03, 0.43], $r^* = 0.28$, $BF_{+0} = 4.18$.

Most critically, we assessed evidence for a relationship between o_v and aMT-birds after controlling for these third variables. We used hierarchical linear regression to ask whether we could predict o_v using aMT-birds with Matrices, P-Speed, and HEVA scores alongside age as nuisance predictors. We use o_v as the criterion here as the Matrices, P-Speed, and HEVA are presented in the visual modality such that modality-specific contributions can be accounted for. We used a model-comparison approach to test whether aMT-birds uniquely improved predictions of o_v. Against a null model with only the nuisance predictors ($R^2 = 0.23$), we found substantial evidence that aMT-birds improved prediction of o_v, $R^2 = 0.36$, $BF_{10} = 27.00$, full model standardized β: aMT-birds = 0.33, 95% CI [0.15, 0.51]; Matrices = 0.17 95% CI [−0.01,

[2] We also repeated our analyses using PCA to aggregate measures and found essentially identical results without requiring the normality assumption in the z-score aggregate.

[3] This is calculated following (Wang & Stanley, 1970) with equal weighting between tests.

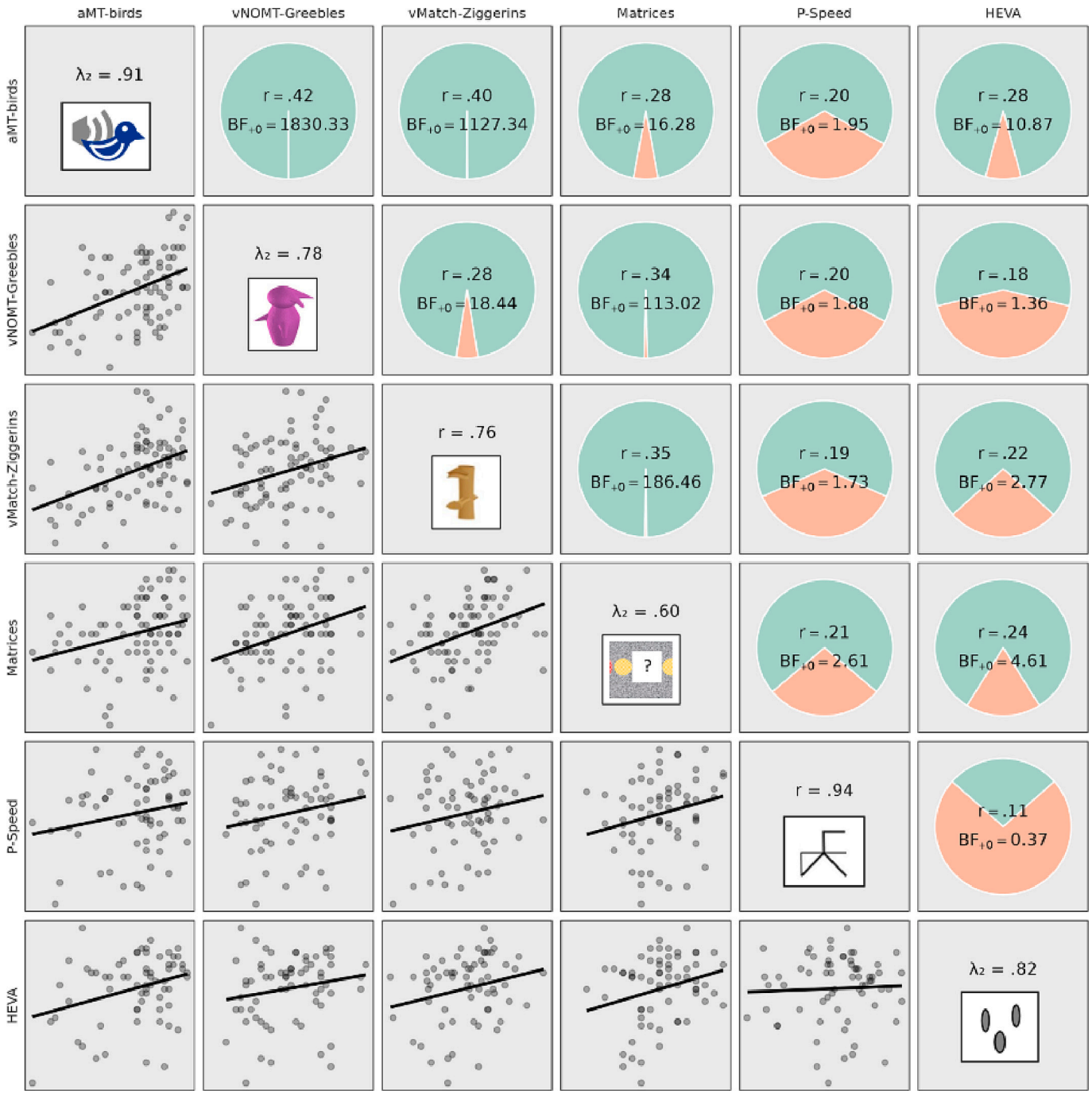

Fig. 2. Zero-order correlations between all tests. The lower triangle are scatterplots for each pair of tests. The upper triangle are the correlation coefficients and directional Bayes factors, the pie charts visualize the relative evidence between the null hypothesis (no correlation; red) and the alternative hypothesis (positive correlation; green). The diagonal indicates unstandardized test reliability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.34]; Age = 0.09, 95% CI [−0.07, 0.27]; P-Speed = 0.12 95% CI [−0.07, 0.29]; HEVA = 0.04, 95% CI [−0.13, 0.20]. This suggests that aMT-birds can uniquely predict visual ability, above and beyond our other covariates. Indeed, even though all 3 control variables were presented in a visual format, aMT-birds was the strongest predictor for o_v with the other predictors only weakly contributing to the best fitting model. To characterize the effect size of the unique relation between aMT-birds and o_v, we controlled for scores on Matrices, P-Speed, and HEVA along with age. The partial correlation between aMT-birds and o_v after controlling for these variables was $r_{xy\bullet z}$ = 0.40, 95% CI [0.20, 0.58], BF_{+0} = 259.24 (Fig. 1).

1.3. Discussion

We found a robust correlation between performance on an auditory object recognition test (aMT-birds) and general visual object recognition ability (o_v). This correlation remained robust even after controlling for potential third variables that include general intelligence, perceptual speed and low-level visual ability. While these results are promising in supporting overlapping mechanisms between auditory object recognition ability and visual object recognition ability, we were only able to measure auditory object recognition ability with one test (see supplemental materials). Just as experimental psychology depends on

converging operations (Garner, Hake, & Eriksen, 1956) to produce strong conclusions, the study of individual differences depends on pooling from different measures to increase construct validity (Rushton, Brainerd, & Pressley, 1983). The correlation we observed could be due to test-specific effects in aMT-birds and therefore, only weakly supports a more general relationship between visual and auditory object recognition ability. For instance, the relationship between aMT-birds and o_v could be driven by memory ability, as aMT-birds and vNOMT-Greebles (a part of o_v) share similar task demands, largely hinged on memory. Even vMatch-Ziggerins, while much less dependent on long-term memory, still depends on working memory, which is known to correlate with long-term memory (Unsworth, 2019). Of course, memory is in some form required for any auditory test, as sounds are sequentially presented. In Experiment 2, we address these concerns in three ways. First, we add two other auditory tests to provide a better estimate of o_a. Second, we include in our measurement of o_v a test that involves simultaneous matching. Third, we measure and control for working memory directly using two tests.

2. Experiment 2

To better measure our variables of interest, we further developed two more auditory object recognition tests to allow us to create an aggregate o_a. We also changed our measures of o_v to the current best practice (Smithson et al., 2023) using the aggregate of three tests, including one that should have minimal reliance on memory and learning ability. As we found that our extra measures of general intelligence, perceptual speed, and low-level visual ability did not strongly account for the relationship between aMT-birds and o_v, we only measured the best covariate of the three (general intelligence). We also included two measures of working memory to test whether it accounts for the correlation across modalities of object recognition.

2.1. Method

2.1.1. Participants

As in Experiment 1, we recruited participants from Prolific and used a Bayesian optional stopping rule to determine our final sample size. In this experiment, we set our minimum sample size at 100 with a target of either $BF_{+0} > 3$ or $BF_{+0} < 1/3$ for all pairwise correlations between object recognition tests within modality. Data collection was split across three sessions to avoid fatigue effects and we only used data from participants who completed all three sessions. Session 1 was approximately 25 min long, session 2 was approximately 20 min long, and session 3 was approximately 27 min long. Participants were paid between 8 and 9 dollars an hour. Participants were invited to each session after they completed the preceding sessions. All participants completed all three sessions within a two-week period. In total, we had 102 participants (39 female, 59 male, 4 other; mean age = 28.62 years, SD = 4.74 years). Data collection and procedures were approved by the Vanderbilt Institutional Review Board.

2.1.2. Test battery

We developed two new auditory object recognition tests using piloting procedures as described in Experiment 1.

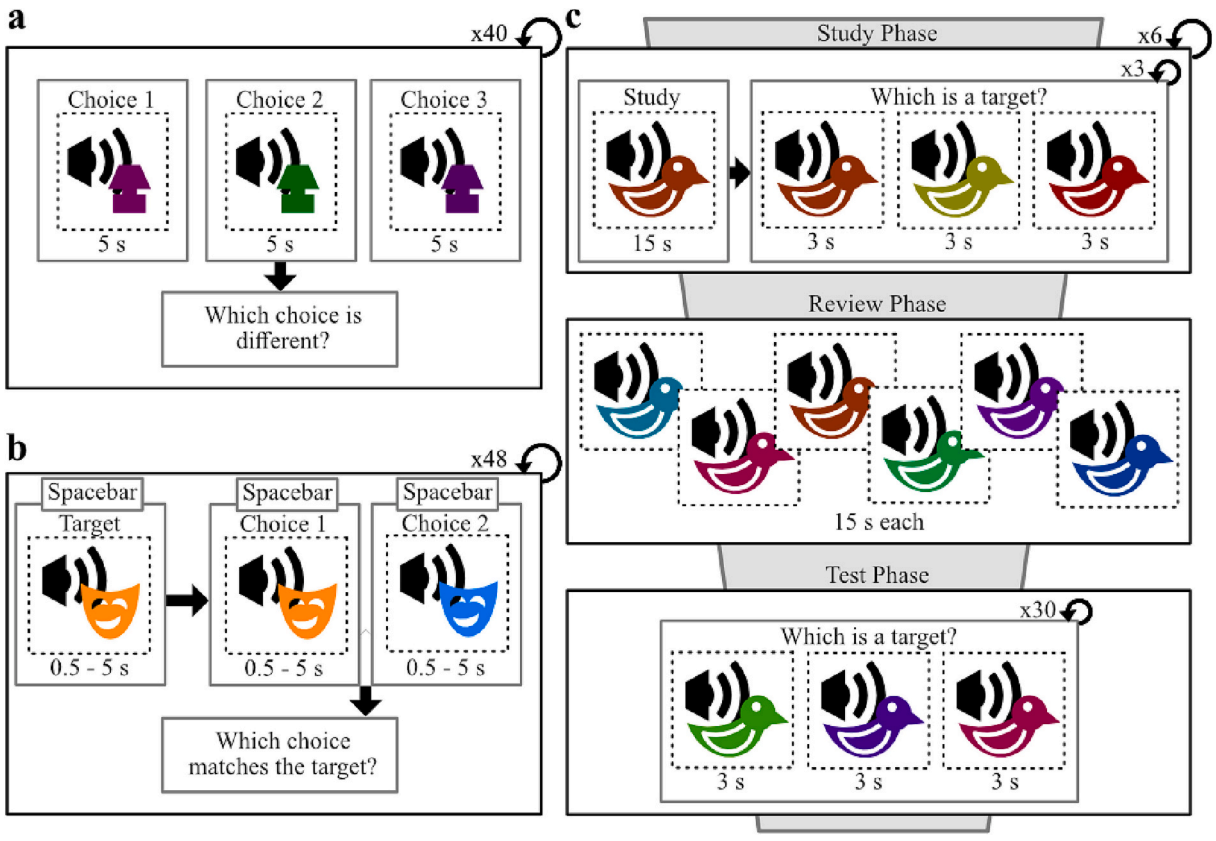

Fig. 3. Schematics of the three auditory object recognition tests. a) aOddball-switches, participants heard three audio clips of keyboard switches where one clip came from a different keyboard switch that they had to choose as the odd one out. The audio clips came from six different switches. b) a2Match-laughs, participants heard an audio clip of a target person laughing, then two other clips of people laughing where they had to choose the clip that is the same person as the target. Some clips were very short, so participants had to initiate each clip with the spacebar. The targets changed every trial. c) aMT-birds, participants studied one of six target birdsongs with easy test trials to select the target, then reviewed each target birdsong, and then performed test trials where any of the six target birdsongs could be used.

2.1.2.1. Auditory Oddball-switches (aOddball-switches; Fig. 3a). In this test, participants had to select the oddball amongst three audio clips of mechanical keyboard switches being pressed. Six mechanical keyboard switches were chosen, such that they differed in timbre and pitch of the noise that was caused when they were pressed. We recorded clips of switches being pressed at varying intervals and pressure. Each clip was edited to be five seconds long with Brownian noise at an amplitude of 0.5 overlaid using Audacity. The added noise increased the confusability of the switches and generally increased difficulty without adding diagnostic information. Each switch had six different clips used throughout the test.

On each trial, three clips were played, one after another, with a one second noise mask distractor of many switches being pressed rapidly between each. Afterwards, participants were prompted to respond which clip came from a different keyboard switch than the other two. The trial ended when the participant made a response, the audio clips did not repeat in a trial. The test trials were ordered by difficulty, from easiest to hardest. The test started with a single practice trial with corrective feedback that repeated until the participant responded correctly. Afterwards, the participant completed 40 trials without feedback. This test was scored with percent correct, therefore, 33% correct was chance performance.

2.1.2.2. Auditory 2AFC matching-laughs (a2Match-laughs; Fig. 3b). In this test, participants listened to a short clip of one person laughing and then had to choose which of two other clips of people laughing was the same person as in the initial clip. We collected audio clips from three different databases (Bachorowski & Owren, 2001; Lavan, Scott, & McGettigan, 2016; Petridis, Martinez, & Pantic, 2013) and edited out non-laughs parts of the clips. The clips varied in length (from half a second to five seconds). Some clips also included irrelevant information (such as recording quality). To ensure that this information was non-diagnostic, individual trials only used clips from the same database.

On each trial, participants heard a target clip of a person laughing followed by two choices, one audio clip of the same person laughing and another of another person laughing. Each clip was initiated by pressing the spacebar to ensure participants were ready to hear the potentially quick audio clip. After the clips, participants were prompted to respond which choice was the same person as in the target clip. Participants completed 48 trials with no practice trials or feedback on any trial. The test was scored with percent correct, therefore, 50% correct was chance performance.

2.1.2.3. Auditory memory test-birds (Fig. 3c). As we found that the aMT-birds performance in Experiment 1 was fairly high (mean accuracy = 72%, SD = 17%), we modified the test by shortening the longer bird clips down to three seconds. Additionally, we no longer repeated the clips during the test phase such that on each trial, each clip was only heard once.

2.1.2.4. Visual object recognition tests. We measured o_v using three visual object recognition tests recently used together to provide a rapid, reliable and valid estimate of o_v (Smithson et al., 2023). They include the vNOMT-Ziggerins (following the same format as the vNOMT-Greebles used in the Experiment 1, with different objects; Ziggerins are synthetic novel objects defined vertical rods with a geometric protrusion at the top and bottom), a modified version of the vMatching test with Greebles and a new Oddball test.

The visual 3AFC Matching-Greebles test (v3Match-Greebles) tasked participants with selecting an object that matches the target on each trial and used asymmetrical Greeble objects. On each trial, participants saw a single target Greeble for 300-1000 ms, followed by a 500 ms visual mask of scrambled Greeble parts with visual noise, and finally three Greebles were presented such that participants would click on the Greeble they believed to be the target. The target presentation time, viewpoint differences, and visual noise on the choices varied across 48 trials to vary difficulty. This test was scored with percent correct, therefore, 33% correct was chance performance.

The visual Oddball-many test (vOddball-many) tasked participants with selecting the different object amongst a set of three objects. This test used objects from a different novel object category on each trial. Each trial began with a fixation for 250 ms, followed by the presentation of three objects for 750-4000 ms and after the objects offset, three blank squares appeared in place of each object, prompting participants to click the square where the object that was different than the other two was. The objects were presented at slightly different sizes and offset vertically to avoid the use of low-level feature comparisons to solve the task. After a response was made, participants were given feedback, though as the object category changed every trial, this feedback did not support learning about the object categories. The presentation time, viewpoint differences and the amount of visual noise varied across 45 trials, as selected based on initial pilot testing to achieve sufficient difficulty. This test was scored with percent correct, therefore, 33% correct was chance performance.

2.1.2.5. Additional tests. In addition to object recognition tests, we also measured general intelligence using the Matrices tests as in Experiment 1, and memory ability using two tests of working memory (Wilhelm et al., 2013). These tasks were used with several others to test hypotheses about working memory at the latent level. They tapped into a common working memory capacity factor suggested to depend on the building, maintaining and updating of arbitrary bindings.

The Binding-Verbal Numerical test tasked participants with remembering word-number pairs. On each trial, participants were presented with a sequence of word-number pairs for 2000 ms each with a 1000 ms gap between each pair. Afterwards, participants were probed with either a word or a number with all the number or words presented during that trial, respectively, as choices. The trial ends when the participant selected a response. Participants performed 17 trials, ranging from two to six word-number pairs. In the final two trials, every word in the trial was probed one at a time with two foils each time from previous trials. A total of 27 responses were used to score the test.

The Operation-Span test tasked participants with remembering series of letters while performing a distraction task judging whether simple arithmetic operations were correct. On each trial, participants were presented with an alternating sequence of letters and arithmetic operations. Only consonants were used for letters and the arithmetic operations were simple addition or subtraction operations with operands ranging between 1 and 10 where half of the arithmetic operations were incorrect. Letters were presented for 1000 ms and the arithmetic operations were presented for 3000 ms or until a response of correct/incorrect was made. Afterwards, participants were prompted to type the exact sequence of the letters during the trial, guessing where they were not sure to preserve the exact letter positioning. Participants performed 15 trials ranging from three to eight letters to remember. Trials were scored based on the proportion of letters correctly remembered in each set, with the test scored as average proportion correct.

2.1.3. Procedures

Data collection was spread over three sessions to avoid fatigue. In the first session, participants completed the aOddball-switches, vNOMT-Ziggerins, and the Matrices-Visuospatial test. In the second session, participants completed aMT-birds, the v3Match-Greebles, and the Matrices-Semantic test. In the third session, participants completed the a2Match-laughs, the vOddball-many, the Binding-Verbal Numeric test, and the Operation-Span test. Each participant completed all three sessions within the span of a week.

2.1.4. Analyses

We only analyzed data from participants who completed all tests

across all three sessions. When calculating aggregate scores for o_a and o_v, if a participant scored below chance on a single test, we did not use that test in creating the aggregate and used the remaining two tests to form the aggregate score for that participant, these procedures were only used for calculating aggregate measures. This avoids concerns of lower than chance performance being due to not understanding the task (instead of having exceptionally low ability, which should still be reflected on the other two measures). A total of 21 participants had an aggregate score that excluded at least one test with less than chance performance. Analysis repeated without this procedure produces similar results. We analyzed our data in a Bayesian framework as in Experiment 1.

2.2. Results

2.2.1. Correlations between object recognition measures

Descriptive statistics for each test and pairwise disattenuated correlations are in Supplementary materials (Table S2). We assessed bivariate correlations for all pairs of object recognition tests in our battery (Fig. 4). In this sample, estimates of reliability for each measure were acceptable. Generally, we found evidence of positive correlations between our auditory object recognition tests. The exception was the correlation between aMT-birds and a2Match-laughs, though the credible intervals and the disattenuated correlations suggest that the true effect size may be similar to the other correlations, $r = 0.15$, 95% CI [0.00, 0.29], $r^* = 0.20$. We replicated the expected positive correlations between the

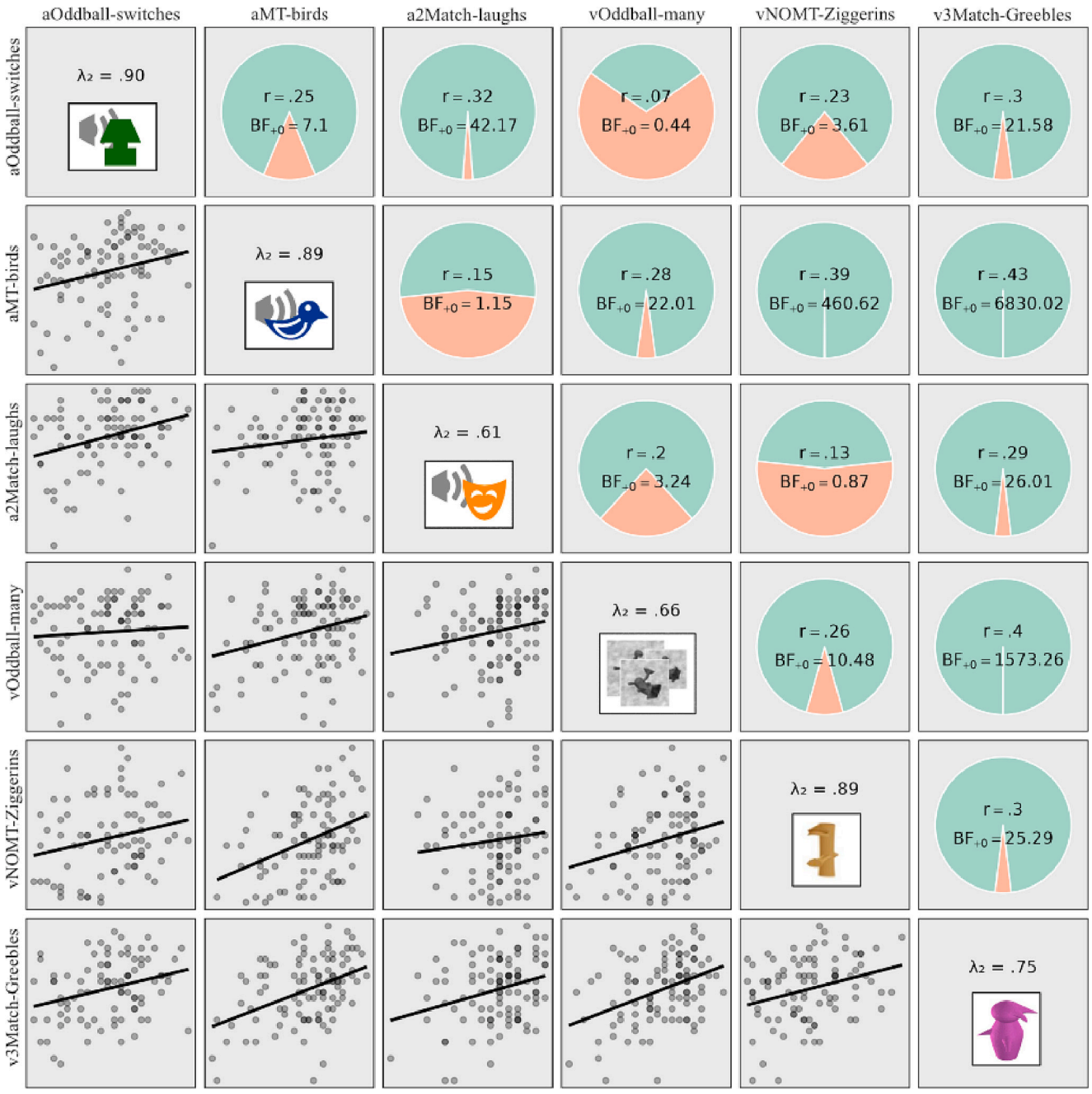

Fig. 4. Zero-order correlations and reliability. The lower triangle are scatterplots for each pair of tests. The upper triangle are the correlation coefficients and directional Bayes factors, the pie charts visualize the relative evidence between the null hypothesis (no correlation; red) and the alternative hypothesis (positive correlation; green). The diagonal indicates unstandardized test reliability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

visual tests. Interestingly, across modality, we found at least substantial evidence for positive correlations between pairs of tests (Fig. 4, top right cells), which suggests that there may be shared factors not only within but also across modalities.

It is important to acknowledge that we do not test whether auditory tests or visual tests correlate with each other more strongly than with those of the other modality (or even between the object recognition tests and other measures), as this is not a prediction we made. For instance, if there is no shared variance based on modality, tests could all equally tap into an amodal object recognition ability. In what follows, we aggregated the tests into separate scores for o_v and for o_a based on their format and estimated the variance for visual and for auditory abilities, setting aside specific task constraints, before we assessed the shared variance between the two abilities.

Both aggregate measures demonstrated high reliability in our sample, o_a aggregate reliability = 0.87; o_v aggregate reliability = 0.86. We found decisive evidence for a positive correlation between o_a and o_v, r = 0.51, 95% CI [0.34, 0.63], r* = 0.56, BF_{+0} = 706,897.70. This provides evidence of overlapping mechanisms in object recognition ability across modalities.

2.2.2. Accounting for non-o variance

As in the previous experiment, we wanted to test whether the correlation between o_a and o_v can be explained by other domain-general abilities: general intelligence (as measured by our two Matrices tests; aggregate reliability = 0.76) and working memory (as measured by two tests; aggregate reliability = 0.82). We found decisive evidence that o_a correlates with general intelligence, r = 0.36, 95% CI [0.18, 0.51], r* = 0.42, BF_{+0} = 367.06, and working memory, r = 0.42, 95% CI [0.23, 0.55], r* = 0.46, BF_{+0} = 3551.48. Similarly, we found decisive evidence that o_v correlates with general intelligence, r = 0.39, 95% CI [0.20, 0.54], r* = 0.45, BF_{+0} = 959.82, and working memory, r = 0.42, 95% CI [0.23, 0.55], r* = 0.47, BF_{+0} = 3591.56.

We used hierarchical linear regression to test whether o_a predicted o_v above and beyond the other covariates. Again, using a model comparison approach, we defined a null model with general intelligence and working memory predicting o_v. We compared this to an alternate model with the addition of o_a as a predictor alongside the other covariates. Against the null model (R^2 = 0.24), we found strong evidence that o_a improved prediction of o_v, R^2 = 0.34, BF_{10} = 37.14, full model standardized β's: o_a = 0.36, 95% CI [0.17, 0.53]; general intelligence = 0.19,

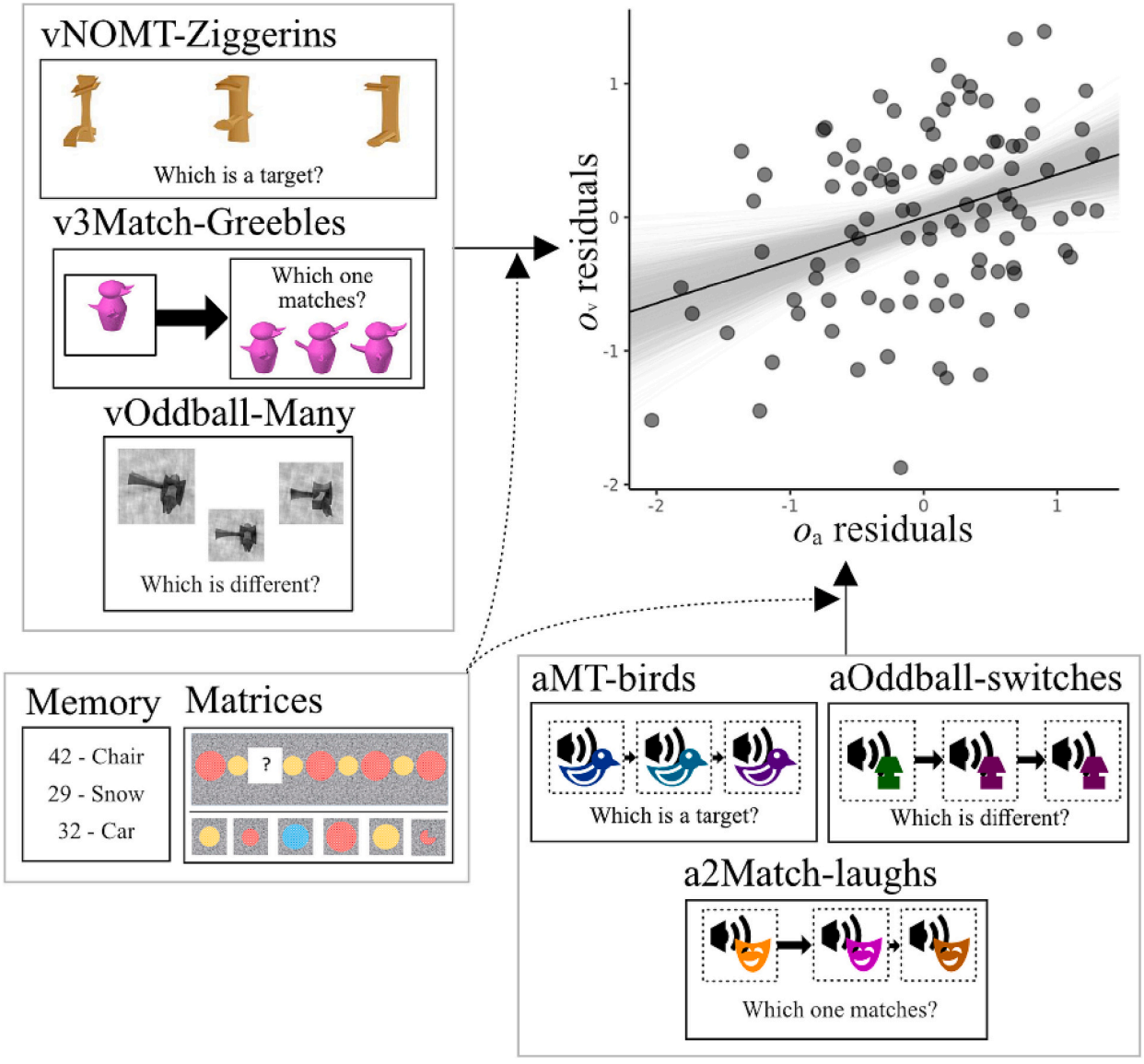

Fig. 5. Partial correlation between o_a and o_v controlling for performance in Matrices and working memory ability. Each point in the scatterplot is a participant, the solid black line is a fit line with light gray lines representing different draws from the posterior.

95% CI [0.01, 0.37]; working memory = 0.20, 95% CI [0.01, 0.39]. This suggests that o_a can predict o_v, above and beyond general intelligence and working memory. This is despite covariates being measured in the visual modality, just like the criterion o_v. To calculate the effect size of the relation between o_a and o_v, controlling for general intelligence and working memory, we computed the partial correlation. We found decisive evidence for a correlation between o_v and o_a, even after accounting for general intelligence and working memory, $r_{xy \bullet z} = 0.33$, 95% CI [0.16, 0.50], $BF_{+0} = 264.99$ (Fig. 5).

2.2.3. Does o have a hierarchical structure?

When we started this project, we would not have been surprised if o_a and o_v had been only minimally related, because auditory perception differs in several ways from visual perception. However, given the results of Experiments 1 and 2, producing clear support for a robust correlation between o_a and o_v, it is important to consider what this means. Our analyses so far have provided strong evidence for shared variance between o_a and o_v, suggesting an amodal *o* (we cannot distinguish between bimodal vs. amodal constructs here, but given our recent work finding overlap between o_h and o_v, we choose the term amodal; see Chow et al., 2023).

In contrast, our analyses so far *do not* provide direct evidence for o_a and o_v as separate abilities. Perhaps the most famous hierarchical theory of abilities is the 3-stratum theory of intelligence (Carroll, 1993) with *g* at the top of a pyramid, contributing to several broad abilities (e.g., fluid reasoning or auditory processing), each of which in turn contributes to more narrow abilities. Likewise, one could conceive of an amodal object recognition ability that contributes to distinct (but correlated) abilities for each modality. Setting aside the challenge of deciding whether an amodal *o* is distinct from other domain-general abilities, this question comes down to whether there is non-error variance shared between indicators of o_a that is also not correlated with o_v. For example, in our work measuring o_v and o_h (with haptic tests, Chow et al., in review), we found evidence for an amodal *o* (a correlation of ~0.5 between o_a and o_h at the latent variable level) as well as for separate o_v and o_h (leaving approximately 75% of the variance to be explained at a modality-specific level).

What is the evidence for separate o_a and o_v in our data? To address this question more directly, we used confirmatory factor analysis (CFA; using lavaan; Rosseel, 2012) to compare a single factor (*o*) model that contributes to all object recognition tasks, $\chi2(12) = 7.71$, $p = 0.56$, RMSEA = 0.00, AIC = −763.84, BIC = −735.11, against a two factor model where o_v and o_a are separate factors that contribute to the tasks in the corresponding modality, $\chi2(13) = 7.65$, $p = 0.47$, RMSEA = 0.00, AIC = −761.90, BIC = −730.77. While both models fit well, the single factor model fit better, despite being the simpler model. Further, the correlation between o_a and o_v in the two-factor model was very high, $r = 0.96$. This provides little support for separate o_a and o_v.

We can extend our CFA into structural equation modeling to ask whether the very strong correlation between o_v and o_a is accounted for by working memory and general intelligence covariates. In order to fit this model, we used the parceling method (Little, Rhemtulla, Gibson, & Schoemann, 2013) to create four indicators for each of these factors (see supplementary information for details). This approach is like the regression approach used in the previous section, but instead of assuming perfect measures, it accounts for measurement error. The results of this analysis support the conclusion that the strong correlation between o_v and o_a (now at the latent level) remains robust ($r = 0.92$) even when controlling for working memory and general intelligence.

Interestingly, exploratory analyses suggest that some of the tasks we used depend mainly on an amodal *o*, while others may also tap an additional modal influence. To look at this, within each modality, we conducted multiple regressions in which we predicted the aggregate of two tasks with the third one from the same modality, controlling for the aggregate of all 3 tasks from the other modality. If a task is a significant predictor of the other two within the same modality, after controlling for the other *o*, it suggests that it taps into a modality-specific source of variation. We found evidence of a modality-specific effects on every task (Table 1) except for aMT-birds and a2Match-laughs. This suggests that, despite a higher order amodal ability, some tests also tap into a modality-specific sources of variation. Taken together with our CFA results, this emphasizes the importance of measuring *o* with multiple tests to estimate a higher-order general object recognition ability. These analyses were not planned and future work with a larger number of tasks – tapping into diverse task demands – would be necessary to better characterize how specific measures tap amodal vs. modality-specific abilities.

2.3. Discussion

With three auditory object recognition ability tests, we measured a domain general auditory object recognition ability analogous to how it is measured in vision. Extending our findings in Experiment 1, we found a robust correlation between o_a and o_v using both the aggregate method and CFA, suggesting overlapping mechanisms across modalities. This correlation is so strong that our results provide much clearer evidence for an amodal *o* than for separate o_a and o_v constructs. Importantly, the shared variance between these visual and object recognition tasks does not appear to be due to general intelligence or working memory. These results can be summarized by suggesting that whatever is common across our visual tasks is the same ability that is common across our auditory tasks, and yet is distinct from memory ability and general intelligence. Because this is the very first study comparing visual and auditory object recognition ability, our conclusions are highly dependent on the specific tasks we used. While latent variables represent abstract constructs, they still are defined by the group of indicators they are based on. As the study of these abilities continues, we can learn which of our conclusions are independent of the exact tasks used to define the constructs.

3. General discussion

With a new battery of high-level auditory tasks, we measured general object recognition ability in the auditory modality for the first time. Like the visual tests used to measure *o* in vision, our auditory tests each measured the ability to match and/or memorize confusable items within a homogenous category. In Experiment 2, we purposely used three auditory tests that vary in their task demands and in their stimulus content, with no overlap in content with our visual tests. The tests were reliable and because of the task-specific differences, their correlation supported a general object recognition ability that is used in the auditory modality, which had not been previously described.

A correlation between auditory and visual performance is not unique in the literature, but our findings are unique because they reflect *domain-general* abilities. Domain-general visual or auditory abilities are not specific to a certain domain, stimulus class, or task demands, and for this reason, they can be expected to generalize more broadly. In contrast, many of the past studies that report correlations between auditory and

Table 1
Partial correlation between a test and o in the same modality, controlling for the other o.

Test	$r_{xy \bullet z}$	95% CI	BF_{+0}
aOddball-switches	0.27	[0.08, 0.44]	17.59
aMT-birds	0.11	[0.00, 0.23]	0.43
a2Match-laughs	0.15	[0.00, 0.30]	1.18
vOddball-many	0.32	[0.16, 0.50]	161.25
vNOMT-Ziggerins	0.22	[0.04, 0.39]	4.93
v3Match-Greebles	0.29	[0.11, 0.46]	53.58

Note. The *o* in the same modality is the aggregate of the remaining two tests within the same modality whereas the other *o* is the aggregate of all tests in the other modality.

visual processing have done so within a specific domain. For instance, in speech studies, a visual-composite score was correlated with an auditory composite score (Watson et al., 1996). In another study, auditory and visual consonant recognition scores were correlated (Grant & Seitz, 1998). In another domain, meta-analysis of research on "sensory sensitivity" in the context of food (generally sensory thresholds) found a correlation between auditory and visual sensitivities (Ginieis, Abeywickrema, Oey, Keast, & Peng, 2022); this was supported in a follow-up experiment where food-relevant visual and auditory tests correlated. Critically, in both the speech and food studies, a correlation between auditory and visual performance could be attributed, in whole or partially, to domain-specific experience.

Admittedly, other correlations between auditory and visual performance may be more difficult to attribute to experience, or at least not domain-specific experience. For instance, in a small study with people varying in their musical expertise, identifying common objects in noise was correlated with recognizing speech in noise (Anaya, Pisoni, & Kronenberger, 2016). Studies measuring the threshold inspection time required for simple judgments found a correlation between visual and auditory tests (e.g., line length or pitch discrimination, Deary, Caryl, Egan, & Wight, 1989). Authors have proposed this may be due to nerve conduction speed (Reed, 1988) or neural efficiency (Hendrickson & Hendrickson, 1980). Simple inspection times are also related to IQ (Deary et al., 1989). Critically, our results depart from these general effects because we found that the correlation between o_v and o_a remains robust even when controlling for intelligence and working memory.

While the logic of our study required that we define latent variables for o_a and o_v separately, our results found little support for the presence of two separate constructs. This is surprising, because we expected auditory object recognition to be at least as different from visual object recognition as we recently found for haptic object recognition (the correlation between o_v and o_h was ~0.5 in recent work, Chow et al., submitted). One reason for this difference could be that o_v in the study on haptic ability was measured with only two tasks, a memory and a matching task similar to those we used here. The more diverse the set of indicators used to define a construct, the more the construct departs from the specific task demands to capture something more general. Therefore, the o_v measured in the present work may be somewhat more general than that measured in the haptic study, and as a result it may be more similar to o_a. Additional work is required to measure object recognition abilities with visual, haptic, and auditory indicators to assess whether visual and auditory abilities are really more related than visual and haptic ones, and whether a hierarchical model can be rejected. The present results are particularly surprising because while we mostly found evidence for an amodal *o,* we found this general ability to be distinct from intelligence, working memory, perceptual speed or early visual abilities.

Our work does not directly specify a mechanism for an amodal *o*, but it provides some clues. Theoretically, *o* resembles acuity, the ability to distinguish details of objects at a distance, but it reflects higher-level perception than the discriminations along single dimensions (e.g., as in the HEVA battery). The neural correlates of o_v are found throughout areas in the ventral occipital cortex – in these regions, those with a higher o_v show greater release of fMRI-adaptation in response to small differences in the shape of novel objects (McGugin, Sunday, & Gauthier, 2022). This occurs even though these differences were task-irrelevant during the scan and the novel objects were all from different categories (as in our vOddball task). Accordingly, objects varying along multiple dimensions are encoded with more precision by those with a higher o_v, before even knowing which of these dimensions is relevant. It is possible that those with a high amodal *o* are likewise better at representing multiple features of a complex auditory object, preparing them better for a variety of subsequent tasks. While we found that working memory broadly defined does not capture this variability, a relevant construct may be the precision of working memory. That is, while working memory capacity is usually measured for items that are easily distinguished, the precision of working memory can be measured with more confusable items. The precision of encoding into working memory is affected by experience with complex stimuli, without changing estimates of the capacity of working memory per se (Lorenc, Pratte, Angeloni, & Tong, 2014; Scolari, Vogel, & Awh, 2008). While the capacity of working memory is related to intelligence, the precision of the encoded representations may not be (Fukuda, Vogel, Mayr, & Awh, 2010). At least in the absence of large differences in experience, *o* may capture individual differences in the precision of encoding in working memory. Our results suggest that some of the underlying mechanisms may be amodal.

Neuroimaging studies have identified brain areas that act as bimodal or even trimodal convergence zones (Kassuba, Menz, Röder, & Siebner, 2013; Man, Damasio, Meyer, & Kaplan, 2015; Porada, Regenbogen, Freiherr, Seubert, & Lundström, 2021). These results are relatively domain specific, such as the planum temporale for audiovisual integration for piano playing (Hasegawa et al., 2004) or the anterior inferior parietal lobe for audiovisual integration of tool perception (Kassuba, Pinsk, & Kastner, 2020). More general work (using a range of toys, tools, and musical instruments) identified the left fusiform gyrus as responding more to objects than textures in the visual, auditory, and tactile modalities (Kassuba et al., 2011). An alternative explanation is the idea of the brain as metamodal (Pascual-Leone & Hamilton, 2001). By this model, brain areas are involved in specific types of information processing, and as they compete to perform tasks, they exhibit modal responses because some sensory inputs are more helpful for a given computation. One could ask why neuroimaging work on the neural substrates of o_v did not identify responses in auditory cortex if o_v and o_a are so strongly correlated? This is likely because in that study (McGugin et al., 2022), sensitivity to visual shape was measured. We would expect that *o* would likewise predict the neural sensitivity to complex sounds (e.g., in an auditory fMRI-adaptation paradigm, Altmann, Doehrmann, & Kaiser, 2007) in the superior temporal gyrus.

We measured visual and auditory abilities separately and did not consider multimodal integration. Studies of multisensory interactions reveal a host of heteromodal subcortical (superior colliculus) and cortical areas (in temporal, parietal and frontal cortex) that may support integration (Macaluso & Driver, 2005). Audiovisual integration can occur very early in processing in visual cortex (Giard & Peronnet, 1999), with neural oscillations supporting cross-modal influences (Bauer, Debener, & Nobre, 2020). Because the brain develops in a multisensory world, individual differences in perception could be influenced by integrative processes. Within a specific domain like musical training, complex multisensory experiences can influence multisensory binding (Lee & Noppeney, 2011) but may also influence domain-general unimodal performance (Habibi, Damasio, Ilari, Elliott Sachs, & Damasio, 2018). Neurodevelopmental variability (e.g., in the degree of pruning of synaptic connections supporting sensory integrations) is one mechanism that could drive individual differences (Ward, 2013) as metamodal brains develop in a multimodal world. Multimodal interactions may influence performance, as suggested by synesthetes showing enhanced perceptual discrimination in affected modalities, even for unimodal stimuli that do not evoke a synaesthetic percept (Banissy, Walsh, & Ward, 2009).

Domain-general abilities are useful because they can predict behavior in a range of situations. o_v can predict a whole host of visual behaviors (Chang & Gauthier, 2021, 2022; Sunday et al., 2018, 2022) in much the same way that domain-general abilities like intelligence, memory capacity or perceptual speed can predict a wide variety of other behaviors. Therefore, it is particularly interesting that *o,* despite capturing variance that is largely amodal in the present study, is distinct from these well studied general abilities like intelligence. If *o* is, as we propose here, distinct from both low-level perception and high-level cognitive abilities, it has the potential to add significant incremental predictive power in a wide range of situations. This ability would be relevant in any situation where perceptual judgments require encoding

complex stimuli with precision and identifying which subset of dimensions is relevant to the current task. The real world, including most demanding occupations and many of the hobbies people are passionate about, is full of such situations.

CRediT authorship contribution statement

Jason K. Chow: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Thomas J. Palmeri:** Writing – review & editing, Supervision. **Graham Pluck:** Resources, Writing – review & editing. **Isabel Gauthier:** Conceptualization, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition.

Data availability

Deidentified data is available on OSF https://doi.org/10.17605/OSF.IO/4YCQZ

Acknowledgments

The authors thanks Cameron Stockwell for help with data collection and recruitment. We thank Jo-Anne Bachowroski and Nadine Lavan for sharing laugh sounds. This work was supported by the David K. Wilson Chair Research Fund (Vanderbilt University).

Appendix A. Supplementary data

Data and analyses scripts are available on OSF, https://doi.org/10.17605/OSF.IO/4YCQZ. Additional supplementary information is available in the supplementary materials.

References

Albright, T. D. (2013). High-level visual processing: Cognitive influences. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbuam, & A. J. Hudspeth (Eds.), *Principles of neural science* (pp. 621–637).

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex, 17*(11), 2601–2608. https://doi.org/10.1093/cercor/bhl167

Amedi, A. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12*(11), 1202–1212. https://doi.org/10.1093/cercor/12.11.1202

Anaya, E. M., Pisoni, D. B., & Kronenberger, W. G. (2016). Long-term musical experience and auditory and visual perceptual abilities under adverse conditions. *The Journal of the Acoustical Society of America, 140*(3), 2074–2081. https://doi.org/10.1121/1.4962628

Bachorowski, J.-A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science, 12*(3), 252–257. https://doi.org/10.1111/1467-9280.00346

Banissy, M. J., Walsh, V., & Ward, J. (2009). Enhanced sensory perception in synaesthesia. *Experimental Brain Research, 196*(4), 565–571. https://doi.org/10.1007/s00221-009-1888-0

Bauer, A.-K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of neural oscillations and cross-modal influences. *Trends in Cognitive Sciences, 24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press.

Ćepulić, D.-B., Wilhelm, O., Sommer, W., & Hildebrandt, A. (2018). All categories are equal, but some categories are more equal than others: The psychometric structure of object and face cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1254–1268. https://doi.org/10.1037/xlm0000511

Chang, T.-Y., & Gauthier, I. (2021). Domain-specific and domain-general contributions to reading musical notation. *Attention, Perception, & Psychophysics, 83*(7), 2983–2994. https://doi.org/10.3758/s13414-021-02349-3

Chang, T.-Y., & Gauthier, I. (2022). Domain-general ability underlies complex object ensemble processing. *Journal of Experimental Psychology: General, 151*(4), 966–972. https://doi.org/10.1037/xge0001110

Chow, J. K., Palmeri, T. J., & Gauthier, I. (2022). Haptic object recognition based on shape relates to visual object recognition ability. *Psychological Research, 86*(4), 1262–1273. https://doi.org/10.1007/s00426-021-01560-z

Chow, J. K., Palmeri, T. J., & Gauthier, I. (2023). *Distinct but related abilities for visual and haptic object recognition. [Manuscript in review]*.

Deary, I. J., Caryl, P. G., Egan, V., & Wight, D. (1989). Visual and auditory inspection time: Their interrelationship and correlations with IQ in high ability subjects. *Personality and Individual Differences, 10*(5), 525–533. https://doi.org/10.1016/0191-8869(89)90034-2

Dunnette, M. D. (1966). *Personnel selection and placement.* Wadsworth Publishing.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests.* Educational Testing Service.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*(5), 673–679. https://doi.org/10.3758/17.5.673

Gaissert, N., Wallraven, C., Bülthoff, H. H., & Bulthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision, 10*(11), 2. https://doi.org/10.1167/10.11.2

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review, 63*(3), 149–159. https://doi.org/10.1037/h0042992

Gauthier, I., & Fiestan, G. (2023). Food neophobia predicts visual ability in the recognition of prepared food, beyond domain-general factors. *Food Quality and Preference, 103*, Article 104702. https://doi.org/10.1016/j.foodqual.2022.104702

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. https://doi.org/10.1162/089892999563544

Ginieis, R., Abeywickrema, S., Oey, I., Keast, R. S. J., & Peng, M. (2022). Searching for individual multi-sensory fingerprints and their links with adiposity – New insights from meta-analyses and empirical data. *Food Quality and Preference, 99*, Article 104574. https://doi.org/10.1016/j.foodqual.2022.104574

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America, 104*(4), 2438–2450. https://doi.org/10.1121/1.423751

Growns, B., Dunn, J. D., Mattijssen, E. J. A. T., Quigley-McBride, A., & Towler, A. (2022). Match me if you can: Evidence for a domain-general visual comparison ability. *Psychonomic Bulletin & Review, 29*(3), 866–881. https://doi.org/10.3758/s13423-021-02044-2

Habibi, A., Damasio, A., Ilari, B., Elliott Sachs, M., & Damasio, H. (2018). Music training and child development: A review of recent findings from a longitudinal study. *Annals of the New York Academy of Sciences, 1423*(1), 73–81. https://doi.org/10.1111/nyas.13606

Hasegawa, T., Matsuki, K.-I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., & Sadato, N. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research, 20*(3), 510–518. https://doi.org/10.1016/j.cogbrainres.2004.04.005

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*(1), 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Hendrickson, D. E., & Hendrickson, A. E. (1980). The biological basis of individual differences in intelligence. *Personality and Individual Differences, 1*(1), 3–33. https://doi.org/10.1016/0191-8869(80)90003-3

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition, 8*(3–5), 489–517. https://doi.org/10.1080/13506280143000214

Jeffreys, H. (1961). *Theory of Probability* (3rd Edition). Oxford University Press.

Kassuba, T., Klinge, C., Hölig, C., Menz, M. M., Ptito, M., Röder, B., & Siebner, H. R. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *NeuroImage, 56*(3), 1566–1577. https://doi.org/10.1016/j.neuroimage.2011.02.032

Kassuba, T., Menz, M. M., Röder, B., & Siebner, H. R. (2013). Multisensory interactions between auditory and haptic object recognition. *Cerebral Cortex, 23*(5), 1097–1107. https://doi.org/10.1093/cercor/bhs076

Kassuba, T., Pinsk, M. A., & Kastner, S. (2020). Distinct auditory and visual tool regions with multisensory response properties in human parietal cortex. *Progress in Neurobiology, 195*, Article 101889. https://doi.org/10.1016/j.pneurobio.2020.101889

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America, 122*(1), 418–435. https://doi.org/10.1121/1.2743154

Kieseler, M.-L., Dickstein, A., Krafian, A., Li, C., & Duchaine, B. (2022). HEVA – A new basic visual processing test. In *Vision Sciences Society Conference*.

Klatzky, R. L., Lederman, S. J., & Metzger, V. A. (1985). Identifying objects by touch: An "expert system". *Perception & Psychophysics, 37*(4), 299–302. https://doi.org/10.3758/BF03211351

Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General, 116*(4), 356–369. https://doi.org/10.1037/0096-3445.116.4.356

Lacey, S., & Campbell, C. (2006). Mental representation in visual/haptic crossmodal memory: Evidence from interference effects. *Quarterly Journal of Experimental Psychology, 59*(2), 361–376. https://doi.org/10.1080/17470210500173232

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Impaired generalization of speaker identity in the perception of familiar and unfamiliar voices. *Journal of Experimental Psychology: General, 145*(12), 1604–1614. https://doi.org/10.1037/xge0000223

Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences, 108*(51). https://doi.org/10.1073/pnas.1115267108

Lee Masson, H., Bulthé, J., op de Beeck, H. P., & Wallraven, C. (2016). Visual and haptic shape processing in the human brain: Unisensory processing, multisensory convergence, and top-down influences. *Cerebral Cortex, 26*(8), 3402–3412. https://doi.org/10.1093/cercor/bhv170

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*(3), 285–300. https://doi.org/10.1037/a0033266

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics, 76*(7), 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "sinking shafts at a few critical points". *Annual Review of Psychology, 51*(1), 405–444. https://doi.org/10.1146/annurev.psych.51.1.405

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences, 28*(5), 264–271. https://doi.org/10.1016/j.tins.2005.03.008

Man, K., Damasio, A., Meyer, K., & Kaplan, J. T. (2015). Convergent and invariant object representations for sight, sound, and touch. *Human Brain Mapping, 36*(9), 3629–3640. https://doi.org/10.1002/hbm.22867

McGugin, R. W., Sunday, M. A., & Gauthier, I. (2022). The neural correlates of domain-general visual ability. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac342

Morey, R. D., & Rouder, J. N. (2022). BayesFactor: Computation of Bayes factors for common designs. https://CRAN.R-Project.Org/Package=Bayesfactor.

Nunnally, J. (1994). *Psychometric theory* (3E ed.). Tata McGraw-Hill Education.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*(4), 291–303. https://doi.org/10.1038/nrn1364

Pascual-Leone, A., & Hamilton, R. (2001). *Chapter 27: The metamodal organization of the brain* (pp. 427–445). https://doi.org/10.1016/S0079-6123(01)34028-1

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research, 90*, 10–14. https://doi.org/10.1016/j.visres.2013.04.015

Petridis, S., Martinez, B., & Pantic, M. (2013). The MAHNOB laughter database. *Image and Vision Computing, 31*(2), 186–202. https://doi.org/10.1016/j.imavis.2012.08.014

Pluck, G. (2019). Preliminary validation of a free-to-use, brief assessment of adult intelligence for research purposes: The matrix matching test. *Psychological Reports, 122*(2), 709–730. https://doi.org/10.1177/0033294118762589

Porada, D. K., Regenbogen, C., Freiherr, J., Seubert, J., & Lundström, J. N. (2021). Trimodal processing of complex stimuli in inferior parietal cortex is modality-independent. *Cortex, 139*, 198–210. https://doi.org/10.1016/j.cortex.2021.03.008

Reed, T. E. (1988). A neurophysiological basis for the heritability of vertebrate intelligence. In *Intelligence and evolutionary biology* (pp. 429–436). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-70877-0_22.

Richler, J. J., Tomarken, A. J., Sunday, M. A., Vickery, T. J., Ryan, K. F., Floyd, R. J., … Gauthier, I. (2019). Individual differences in object recognition. *Psychological Review, 126*(2), 226–251. https://doi.org/10.1037/rev0000129

Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition, 166*, 42–55. https://doi.org/10.1016/j.cognition.2017.05.019

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2). https://doi.org/10.18637/jss.v048.i02

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*(1), 18–38. https://doi.org/10.1037/0033-2909.94.1.18

Sathian, K., & Lacey, S. (2022). Cross-modal interactions of the tactile system. *Current Directions in Psychological Science*. https://doi.org/10.1177/09637214221101877, 096372142211018.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). The Guildford Press.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15*(1), 215–222. https://doi.org/10.3758/PBR.15.1.215

Smithson, C. J. R., Chow, J. K., Chang, T.-Y., & Gauthier, I. (2023). *Measuring object recognition ability: Reliability, validity and the aggregate z-score approach. [Manuscript submitted for publication]*. https://doi.org/10.31234/osf.io/bmyan

Sunday, M. A., Donnelly, E., & Gauthier, I. (2018). Both fluid intelligence and visual object recognition ability relate to nodule detection in chest radiographs. *Applied Cognitive Psychology, 32*(6), 755–762. https://doi.org/10.1002/acp.3460

Sunday, M. A., Tomarken, A., Cho, S.-J., & Gauthier, I. (2022). Novel and familiar object recognition rely on the same ability. *Journal of Experimental Psychology: General, 151*(3), 676–694. https://doi.org/10.1037/xge0001100

Thaler, L., & Goodale, M. A. (2016). Echolocation in humans: An overview. *WIREs Cognitive Science, 7*(6), 382–393. https://doi.org/10.1002/wcs.1408

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin, 145*(1), 79–139. https://doi.org/10.1037/bul0000176

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40*(5), 663–705. https://doi.org/10.3102/00346543040005663

Ward, J. (2013). Synesthesia. *Annual Review of Psychology, 64*(1), 49–75. https://doi.org/10.1146/annurev-psych-113011-143840

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., & Jensen, J. K. (1982). An auditory discrimination test battery. *The Journal of the Acoustical Society of America, 71*(S1), S73. https://doi.org/10.1121/1.2019532

Watson, C. S., Qiu, W. W., Chamberlain, M. M., & Li, X. (1996). Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *The Journal of the Acoustical Society of America, 100*(2), 1153–1162. https://doi.org/10.1121/1.416300

Wetzels, R., & Wagenmakers, E.-J. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review, 19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00433

Wilmer, J. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision, 21*(6), 561–579. https://doi.org/10.1163/156856808786451408

Wong, A. C.-N. N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – But which type? *Psychological Science, 20*(9), 1108–1117. https://doi.org/10.1111/j.1467-9280.2009.02430.x

Yang, J., Yan, F.-F., Chen, L., Xi, J., Fan, S., Zhang, P., Lu, Z.-L., & Huang, C.-B. (2020). General learning ability in perceptual learning. *Proceedings of the National Academy of Sciences, 117*(32), 19092–19100. https://doi.org/10.1073/pnas.2002903117